

# Multideviations: The hidden structure of Bell's theorems

Brandon Fogel

December 17, 2015

## Abstract

Specification of the strongest possible Bell inequalities for arbitrarily complicated physical scenarios—any number of observers choosing between any number of observables with any number of possible outcomes—is currently an open problem. Here I provide a new set of tools, which I refer to as “multideviations”, for finding and analyzing these inequalities for the fully general case. In Part I, I introduce the multideviation framework and then use it to prove an important theorem: the Bell distributions can be generated from the set of joint distributions over all observables by deeming specific degrees of freedom unobservable. In Part II, I show how the theorem provides a new method for finding tight Bell inequalities. I then specify a set of new tight Bell inequalities for arbitrary event spaces—the “even/odd” inequalities—which have a straightforward interpretation when expressed in terms of multideviations. The even/odd inequalities concern degrees of freedom that are independent of those involved in parameter independence, raising the possibility of a new Bell’s theorem with stronger philosophical implications. Also, contrary to expectations, the violation of the inequalities by quantum mechanics increases in size with the number of systems.

# 1 Introduction

In its original form, Bell’s theorem described a rather simple physical scenario: two observers each choosing between two possible measurements with two possible outcomes each (see Bell 1964, 1966). The derived empirical limits, the Bell inequalities, were eventually given a complete description (see Clauser and Horne 1974; Fine 1982). Subsequent attempts to generalize the theorem to more complicated scenarios have yielded some notable results, but a similarly complete and systematic treatment has not yet been achieved.<sup>1</sup> The primary obstacle has been the computational complexity of the problem, which grows exponentially with each of the parameters, particularly with the number of observers.

The present paper takes a step toward taming that complexity and, in so doing, exposes some of the deeper structure underlying Bell’s theorems.

The paper is organized into two parts. Part I comprises sections 2-5. Section 2 contains non-technical presentations of the main results of Part I. Section 3 introduces a set of new mathematical tools, which I dub “multideviations”: correlation functions that decompose joint probability distributions into independent degrees of freedom for each subset of observers. Because the tools have a generality beyond the application to Bell’s theorem, I provide a systematic, if abbreviated, treatment.

Section 4 shows how to apply the multideviation framework to the distributions used in Bell’s theorems. Section 5 contains an important theorem: the distributions obeying the Bell inequalities are precisely those generated by considering specific multideviation degrees of freedom inaccessible in joint distributions over all observables.

Part II comprises sections 6-10. Section 6 uses the new theorem, along with a bit of matroid theory, a well-established branch of combinatorics, to outline a new method for finding tight Bell inequalities<sup>2</sup> for arbitrary physical

---

<sup>1</sup>Notable early attempts include Svetlichny (1987), Mermin (1990), and Greenberger et al. (1990), which provide derivations of empirical limitations for systems with three or four particles. Peres (1999) provides a general method for deriving inequalities for higher-dimensional systems, although few inequalities are actually presented. Pitowsky and Svozil (2001) provide a complete list of tight Bell inequalities for three observers choosing between two binary observables, and Werner and Wolf (2001) provide a large class of Bell inequalities, many of which are not tight (i.e., maximally restrictive). Other notable partial results include Collins et al. (2002), Żukowski and Brukner (2002), Uffink (2002), and, more recently Bancal et al. 2011.

<sup>2</sup>A tight Bell inequality is a Bell inequality that is extremal in the sense that it cannot

scenarios.

In section 7, I use the new method to provide a large class of solutions for the general case—tight Bell inequalities for any number of observers each choosing between any number of observables each with any number of possible outcomes. In section 8, I provide a convenient conceptual and graphical interpretation of an important subset of the new inequalities, which I refer to as the “even/odd inequalities”. In section 9, I outline an important feature of these inequalities: they concern degrees of freedom that are independent of those involved in parameter independence. This means they could theoretically be derived without that condition or anything equivalent, thus allowing for a sharper philosophical conclusion than is achieved with the full complement of Bell inequalities.

In section 10, I show that quantum mechanics violates a subset of these inequalities with particularly simple form. One somewhat surprising result is that the size of the violation increases with the number of observers and rather quickly converges to the theoretical maximum. If the violation of these inequalities is taken to represent a peculiarly non-classical effect, then one might have expected the effect to diminish as the number of systems is increased, which is often considered a classical limit; yet this is not the case.

## Part I

# Multideviations and the projection theorem

In this part, I present the multideviation framework, show how it is applied to Bell’s theorem, and then prove an important projection theorem concerning distributions satisfying the Bell inequalities. For the reader uninterested in the full technical presentation, I offer in section 2 conceptual presentations of the main results. Section 2.1 describes the new mathematical tools, and section 2.2 describes how these tools provide a unique and fundamental perspective on the mathematical origin of the Bell inequalities.

---

be written as a linear combination of other Bell inequalities.

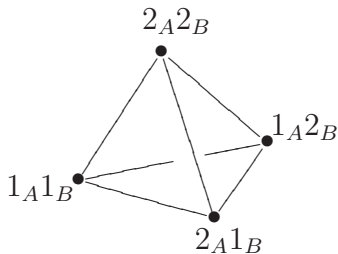


Figure 1: Geometric representation of a set of probability distributions for the 2x2 case. Each vertex represents the distribution where that outcome has probability 1. Each facet represents the distributions where the opposite vertex has probability 0. All other distributions correspond to points inside the tetrahedron.

## 2 Conceptual presentations

### 2.1 Multideviations—A geometric prelude

Consider two observers,  $A$  and  $B$ , each measuring an observable with two possible outcomes, so that there are four possible joint outcomes:  $\{1_A 1_B, 1_A 2_B, 2_A 1_B, 2_A 2_B\}$ . The set of possible probability distributions over these outcomes has three degrees of freedom: one for each outcome minus one for the constraint that the probabilities sum to 1. The set can be represented by a tetrahedron (see fig. 1), where the vertices and facets have convenient interpretations.

The three degrees of freedom in the set of distributions can be broken up in a convenient way. One degree of freedom can be used to describe the probability that  $A$  will observe outcome  $1_A$  or  $2_A$ —the marginal probabilities for  $A$ . Only one degree of freedom is needed because the probabilities for those outcomes must sum to 1. Another degree of freedom can be used to describe the marginal probabilities for  $B$ . These two degrees of freedom are independent of one another; they correspond to different subspaces in the vector space containing the tetrahedron. In fact, the subspaces are orthogonal, and if we project the tetrahedron into the subspace containing both of them, it looks like a square (see fig. 2).

The remaining degree of freedom represents information that is not contained in the marginal probabilities; that is, it concerns only the correlation

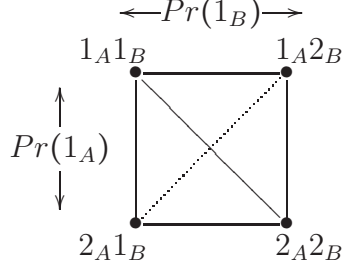


Figure 2: Projection of the set of 2x2 probability distributions into the marginal plane. Four of the tetrahedron's edges project to edges in the square, while two become internal lines (the diagonals). The solid diagonal line represents the closer of those two edges, and the dotted line represents the edge on the opposite side of the tetrahedron.

between their measurements.

Although the three degrees of freedom are linearly independent, they are related to one another by the shape of the tetrahedron. Consider figure 2, which shows the tetrahedron from a particular angle. The center of the tetrahedron projects to the center of the square, between the two edges that project internally (the diagonal lines). If one starts in the center of the tetrahedron and then moves out of the page, one will eventually hit the edge connecting  $1_A 1_B$  and  $2_A 2_B$ . Along this edge,  $Pr(1_A) = Pr(1_B)$  and, consequently,  $Pr(2_A) = Pr(2_B)$ . That is, the correlation degree of freedom is maximized, and the effect is that observers  $A$  and  $B$  will always get the same result. Conversely, if one starts in the center and moves directly into the page, one will hit the edge on the other side of the tetrahedron, where the correlation degree of freedom is minimized and the outcomes are perfectly anti-correlated (this is the geometric interpretation of equation 3.28 in section 3.5.2).

These degrees of freedom correspond to what I have dubbed “multideviations”: special functions that decompose joint probability distributions into linearly independent correlation degrees of freedom. Multideviations are given systematic treatment in section 3. The multideviation decomposition is of fundamental importance to Bell's Theorem (see section 5), and it provides a new method for the determination of Bell inequalities for arbitrary event spaces (see sections 6 and 7).

## 2.2 The projection theorem—a conceptual prelude

The projection theorem of section 5 says that the distributions described by Bell’s theorem can be found by two different methods. The first is familiar: 1) a set of joint measurement contexts is formed by allowing each observer to choose from a set of mutually exclusive observables, 2) a different probability distribution is specified for each joint measurement context, and 3) certain conditions (often of ontological importance) are imposed on them. The second is new: 1) a single joint measurement context is formed by supposing that all observables are measured together, 2) a single probability distribution is specified for that context, and 3) certain degrees of freedom are deemed unobservable. The projection theorem shows that the two methods produce the same distributions.

We will consider the two methods schematically for the simplest case: two observers each choosing between two binary observables. In the first method, observer  $A$  chooses between observables 1 and 2, and observer  $B$  chooses between observables 3 and 4. There are thus four measurement contexts:  $\{13, 23, 14, 24\}$ . We assign a probability distribution,  $P_{ij}$ , to each context, and using the multideviation framework described in the previous section, we can break each distribution into three degrees of freedom:  $Q_{ij}^{\{i,j\}}, Q_{ij}^{\{i\}}, Q_{ij}^{\{j\}}$ . This is depicted in figure 3. At first, there are 12 total degrees of freedom. Parameter independence requires that the marginals for one observer be independent of the choice of observable for the other observer. This is enforced by equating the  $Q_{ij}^\sigma$  for different  $ij$  but the same  $\sigma$ , which removes four degrees of freedom (in the figure, this is indicated by the double arrows). Enforcing determinism or outcome independence for the underlying states is not represented so easily, but neither condition reduces the degrees of freedom of the set of observable distributions (see section 4.2 for more on all of these conditions).

In the second method, we consider all four observables to be measured by independent observers; there are thus four observers,  $\{1, 2, 3, 4\}$ , and a single probability distribution over 16 possible outcomes. There will thus be 16 different multideviation degrees of freedom (Fig. 4).

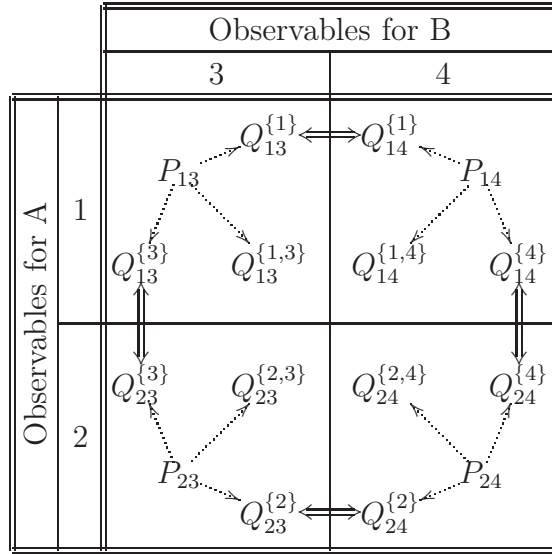


Figure 3: Degrees of freedom for a multiple-context distribution in the 2x2x2 case. There are four measurement contexts and thus four probability distributions. Each distribution is broken into three multideviation degrees of freedom (dotted arrows), and parameter independence requires multideviations of the same order to be the same across measurement contexts (double arrows).

$$\begin{array}{ccccccc}
& & & & Q_{\{1,2,3,4\}} & & \\
& & & & & & \\
& & Q_{\{1,2,3\}} & Q_{\{1,2,4\}} & Q_{\{1,3,4\}} & Q_{\{2,3,4\}} & \\
Q_{\{1,2\}} & Q_{\{1,3\}} & Q_{\{1,4\}} & Q_{\{2,3\}} & Q_{\{2,4\}} & Q_{\{3,4\}} & \\
& Q_{\{1\}} & Q_{\{2\}} & Q_{\{3\}} & Q_{\{4\}} & & \\
& & & & Q_{\emptyset} & & 
\end{array}$$

Figure 4: Multideviation decomposition for 4 observers.

$$\begin{array}{ccccccc}
& & & & \cancel{Q_{\{1,2,3,4\}}} & & \\
& & & & & & \\
& & \cancel{Q_{\{1,2,3\}}} & \cancel{Q_{\{1,2,4\}}} & \cancel{Q_{\{1,3,4\}}} & \cancel{Q_{\{2,3,4\}}} & \\
\cancel{Q_{\{1,2\}}} & Q_{\{1,3\}} & Q_{\{1,4\}} & Q_{\{2,3\}} & Q_{\{2,4\}} & \cancel{Q_{\{3,4\}}} & \\
& Q_{\{1\}} & Q_{\{2\}} & Q_{\{3\}} & Q_{\{4\}} & & 
\end{array}$$

Figure 5: Reduced multideviation decomposition for 4 observers.

One degree of freedom,  $Q_{\emptyset} = \frac{1}{16}$ , is fixed at the outset. The rest can vary between  $\pm Q_{\emptyset}$  and are linearly independent, although they are related to each other via some inequality constraints. Each  $Q$  measures a correlation between a subset of observers that cannot be measured by combinations of the other  $Q^{\sigma}$ .

Suppose now that some of the correlation degrees of freedom are considered unobservable—namely, those in which 1 and 2 are involved in a correlation together or similarly with 3 and 4. There are 7 such degrees of freedom, and, after eliminating those, we are left with 8 (see fig. 5). A careful comparison of the remaining degrees of freedom in figures 3 and 5 shows that there is a 1-1 correspondence.

This correspondence is essentially the content of the projection theorem of section 5: the distributions that obey all Bell inequalities are generated by ignoring multideviation correlations involving mutually exclusive observables in joint distributions on the set of all observables.

### 3 Multideviations

I will now introduce a new framework for analyzing probability distributions; the degrees of freedom of the distributions will be reconfigured in terms of



novel correlation functions that I have named “multideviations”. These do not appear to exist in the current mathematical literature. Although I have devised this framework specifically for studying the distributions in Bell’s theorem, it is very general and may find application elsewhere.

### 3.1 A note on notation

The representation of functions of unknown numbers of variables over products of arbitrary sets can quickly become unwieldy, and the literature on Bell’s theorem has suffered for lack of an efficient and standard notation. For this reason, I have developed a new notational framework, the two basic elements of which are the product set and intuple.<sup>3</sup> A *product set* is the Cartesian product of an indexed family of sets:

$$\Pi A_B \equiv \prod_{i \in B} A_i \quad (3.1)$$

where  $A_B \equiv \{A_i | i \in B\}$ . The elements of product sets are *intuples* (“indexed tuples”) and will be designated by an overhead tilde:  $\tilde{x}_B \in \Pi A_B$ . The intuple components are indicated in straightforward fashion:  $x_i \in A_i$ , where  $i \in B$ . Intuples can be thought of as ordered tuples or, if the components carry the indices of their parent sets, as simple sets.

Given some product set  $\Pi A_B$ , there is a product set  $\Pi A_\sigma$  for every  $\sigma \subseteq B$ . Likewise,  $\tilde{x}_B$  defines an intuple  $\tilde{x}_\sigma$  for every  $\sigma \subseteq B$ .

Summation over product sets can be represented compactly:

$$\sum_{\tilde{x}_\sigma} f(\tilde{x}_B) \equiv \sum_{\underbrace{x_i \in A_i \cdots x_k \in A_k}_{\sigma = \{i, \dots, k\}}} f(\tilde{x}_B) \quad (3.2)$$

where  $\sigma = \{i, j, \dots, k\} \subseteq B$ .

### 3.2 Motivation

Consider a set of observers  $B$ . An observer  $i \in B$  performs a single measurement, where the set of possible outcomes is  $A_i$ . The possible joint outcomes are given by the product set  $\Pi A_B$ , and the probability of getting the joint

---

<sup>3</sup>See Fogel (2011) for a more systematic treatment.

outcome  $\tilde{x}_B$  is given by the ordinary distribution  $P(\tilde{x}_B)$ . Of the many different ways to measure the correlation between outcomes  $x_i$  and  $x_j$  for two different observers  $i, j \in B$ , the most common is the covariance:

$$P^{\{i,j\}}(x_i x_j) - P^{\{i\}}(x_i) P^{\{j\}}(x_j) \quad (3.3)$$

where  $P^\sigma(\tilde{x}_\sigma) \equiv \sum_{B \setminus \sigma} P(\tilde{x}_B)$  is a generalized marginal function.

While the covariance has many useful applications, it has a significant drawback as a pure measure of correlation: it depends on the absolute value of the marginal functions  $P^{\{i\}}(x_i)$  and  $P^{\{j\}}(x_j)$ , not just their relation to one another. For example, the covariance is 0 when  $P^{\{i\}}(x_i) = P^{\{j\}}(x_j) = 1$  and  $P^{\{i\}}(x_i) = P^{\{j\}}(x_j) = 0$ , even though these are states of seemingly high correlation.

It would thus be desirable to have a measure of correlation that is independent of the relevant marginal degrees of freedom. Furthermore, once we have such a measure for the pairs of observers  $\{i, j\}$ ,  $\{i, k\}$ , and  $\{j, k\}$ , we can find a measure of correlation for the triple  $\{i, j, k\}$  that is independent of those as well. Repeating this, we can find an independent correlation function for each  $\sigma \subseteq B$ . It turns out that this demand more or less fixes the form of the functions.

### 3.3 Multideviation seed functions

Given an ordinary distribution  $P(\tilde{x}_B)$  over a product set  $\Pi A_B$ , we begin with an arbitrary linear combination of the elements of the distribution:

$$Q^\sigma(\tilde{x}_\sigma) \equiv \sum_{\tilde{y}_B} q^\sigma(\tilde{x}_\sigma, \tilde{y}_\sigma) P(\tilde{y}) \quad (3.4)$$

We want the marginals to be written as sums of these functions for only the relevant degrees of freedom:

$$P^\sigma(\tilde{y}_\sigma) = c_\sigma \sum_{\rho \in \mathcal{P}(B)} Q^\rho(\tilde{y}_\rho) \quad (3.5)$$

where  $c_\sigma$  is some constant. It turns out that if we demand that  $Q^\sigma(\tilde{x}_\sigma)$  and  $Q^\rho(\tilde{y}_\rho)$  be linearly independent of one another whenever  $\sigma \neq \rho$ , and that  $c_B = 1$ , then the form of the seed function  $q$  is fixed.

**Definition** (Multideviation seed function). Given a product set  $\Pi A_B$  and a cardinality function  $n_\sigma \equiv \prod_{i \in \sigma} |A_i|$ , where  $\sigma \subseteq B$ ,

$$q_B^\sigma(\tilde{x}_\sigma, \tilde{y}_\sigma) \equiv \frac{1}{n_B} \prod_{i \in \sigma} (n_i \delta_{x_i=y_i} - 1) \quad (3.6)$$

is a  $\sigma$ -order multideviation seed function ( $\sigma$ -MSF).<sup>4</sup>

The MSFs reproduce the Kronecker delta:

$$\delta_{\tilde{x}=\tilde{y}} = \sum_{\sigma \in \mathcal{P}(B)} q_B^\sigma(\tilde{x}_\sigma, \tilde{y}_\sigma) \quad (3.7)$$

This means that the MSFs cover the function space; they can be used to decompose any function over the given product set,  $\Pi A_B$ .

The MSFs have some especially useful algebraic properties. They are symmetric in the intuple arguments:

$$q_B^\sigma(\tilde{x}_\sigma, \tilde{y}_\sigma) = q_B^\sigma(\tilde{y}_\sigma, \tilde{x}_\sigma) \quad (3.8)$$

They sum to zero for each argument:

$$\forall i \in \sigma \left[ \sum_{x_i} q_B^\sigma(\tilde{x}_\sigma, \tilde{y}_\sigma) = 0 \right] \quad (3.9)$$

Most important, they are closed and orthogonal under a natural inner product:

$$\sum_{\tilde{y}_B} q_B^\sigma(\tilde{x}_\sigma, \tilde{y}_\sigma) q_B^\rho(\tilde{y}_\rho, \tilde{z}_\rho) = \delta_{\sigma=\rho} q_B^\sigma(\tilde{x}_\sigma, \tilde{z}_\sigma) \quad (3.10)$$

One can think of MSFs as hypermatrices with a particular matrix-multiplication structure. (3.7) says that the MSFs span the entire hypermatrix vector space. (3.9) says that there are linear dependencies among the hypermatrices. (3.10) says that hypermatrices with different  $\sigma$  are orthogonal. Thus, the MSFs can be used to identify a complete set of orthogonal subspaces of the vector space (or, equivalently, the function space) defined by the given product set (see appendix B for more on vector spaces and multideviations).

---

<sup>4</sup>Using a generalization of the binomial theorem, the MSF can be put in an alternate form that is frequently very useful:

$$q^\sigma(\tilde{x}_\sigma, \tilde{y}_\sigma) = \sum_{\mu \in \mathcal{P}(\sigma)} \frac{(-1)^{|\sigma \setminus \mu|}}{n_{V \setminus \mu}} \delta_{\tilde{x}, \tilde{y}}^\mu$$

where  $\delta_{\tilde{x}, \tilde{y}}^\mu \equiv \prod_{i \in \mu} \delta_{x_i=y_i}$ .

## 3.4 Multideviations

### 3.4.1 Arbitrary functions

A multideviation is the portion of a function isolated by an MSF:

**Definition** (Multideviation). Given a field-valued function  $f(\tilde{x}_B)$  on a factorizable set  $\Pi A_B$ , a  $\sigma$ -multideviation is given by

$$Q_f^{\sigma,B}(\tilde{x}_\sigma) \equiv \sum_{\tilde{y}_B} f(\tilde{y}_B) q_B^\sigma(\tilde{x}_\sigma, \tilde{y}_\sigma) \quad (3.11)$$

Reference to the index set,  $B$ , can be omitted if it is clear from the context. Multideviations inherit the algebraic properties of the MSFs. Thanks to (3.7), the multideviations decompose their generating function:

$$f(\tilde{x}_B) = \sum_{\sigma \in \mathcal{P}(B)} Q_f^\sigma(\tilde{x}_\sigma) \quad (3.12)$$

Since the decomposition is invertible, it is also unique. The summation property of the MSFs (3.9) means that the multideviations are not linearly independent within a given  $\sigma$ :

$$\forall i \in V \left[ \sum_{x_i} Q_f^\sigma(\tilde{x}_\sigma) = 0 \right] \quad (3.13)$$

And the inner product property of the MSFs (3.10) means that the multideviations can be picked out by summation with an MSF:

$$\sum_{\tilde{x}_B} Q_f^{\sigma,B}(\tilde{x}_\sigma) q_B^\mu(\tilde{x}_\mu, \tilde{y}_\mu) = \delta_{\sigma=\mu} Q_f^{\sigma,B}(\tilde{y}_\sigma) \quad (3.14)$$

Multideviations provide an alternate representation of a function; they take the function's degrees of freedom and redistribute them.

### 3.4.2 Ordinary probability distributions

There are several advantages to representing ordinary probability distributions in terms of multideviations. One concerns the probability constraint,  $\sum_{\tilde{x}_B} P(\tilde{x}_B) = 1$ . In the natural representation, there is not a particularly easy way to implement this constraint. However, in the multideviation representation, it becomes

$$Q_P^\varnothing = \frac{1}{n_B} \quad (3.15)$$

A single, specific degree of freedom is fixed; the rest are unaffected. A second advantage of the multideviation representation is that the first-order marginal degrees of freedom are isolated from the others:

$$Q_P^{\{i\}}(x_i) = \frac{1}{n_B} (n_{\{i\}} P^{\{i\}}(x_i) - 1) \quad (3.16)$$

The first-order marginals and the first-order multideviations are distinguished only by an offset and a scaling factor; each fixes the other.

The 2nd-order multideviation,  $Q_P^{\{i,j\}}(\tilde{x}_{\{i,j\}})$ , is what is left of  $P^{\{i,j\}}(\tilde{x}_{\{i,j\}})$  after the linear dependences on  $Q_P^{\{i\}}(x_i)$  and  $Q_P^{\{j\}}(x_j)$  have been removed. In terms of  $P$ , this is written most economically as

$$Q_P^{\{i,j\}}(\tilde{x}_{\{i,j\}}) = \frac{1}{n_B} (n_{\{i,j\}} P^{\{i,j\}}(\tilde{x}_{\{i,j\}}) - n_{\{i\}} P^{\{i\}}(x_i) - n_{\{j\}} P^{\{j\}}(x_j) + 1) \quad (3.17)$$

If one is not convinced by (3.10) that  $Q_P^{\{i,j\}}(\tilde{x}_{\{i,j\}})$  and  $Q_P^{\{i\}}(x_i)$  are linearly independent, one need only see how to modify one without the other.<sup>5</sup>

The higher-order multideviations have a similar form:

$$Q_P^\sigma(\tilde{x}_\sigma) = \frac{1}{n_B} \sum_{\rho \in \mathcal{P}(\sigma)} (-1)^{|\sigma \setminus \rho|} n_\rho P^\rho(\tilde{x}_\rho) \quad (3.18)$$

One of the novelties with multideviations is that one can consider correlations between different subsets of observers independently. For example, it is possible for observers 1 and 2 to have outcomes that are highly correlated while observers 1, 2, and 3 collectively do not (and vice-versa).

### 3.4.3 Inequality constraints

The probability axiom,  $P(\tilde{x}_B) \geq 0$ , causes multideviations of different orders to be related by inequality constraints:

$$\sum_{\sigma \in \mathcal{P}(B)} Q_P^\sigma(\tilde{x}_\sigma) \geq 0 \quad (3.19)$$

---

<sup>5</sup>The substitution  $P^{\{i,j\}}(x_i y_j) \rightarrow P^{\{i,j\}}(x_i y_j) + \frac{a}{n_j}$  for all  $y_j$  modifies  $Q_P^{\{i\}}(x_i)$  while leaving  $Q_P^{\{i,j\}}(\tilde{x}_{\{i,j\}})$  unchanged. The substitution  $P^{\{i,j\}}(\tilde{y}_{\{i,j\}}) \rightarrow P^{\{i,j\}}(\tilde{y}_{\{i,j\}}) + a(-1)^{\delta_{y_i=x_i}}(-1)^{\delta_{y_j=x_j}}$  for all  $y_i$  and  $y_j$  does the opposite.

for all  $\tilde{x}_\sigma \in \Pi A_B$ . Because the multideviations are related by (3.13), and because many of them appear in more than one inequality, the overall structure of these inequalities can be rather complicated. The binary case (i.e.,  $|A_i| = 2$  for all  $i$ ) is significantly simpler than the general case (see section 3.5.2).

In general, the inequalities restrict the multideviations to the following range:

$$-\frac{1}{n_{\min} - 1} n_\sigma^* Q_P^\emptyset \leq Q_P^\sigma(\tilde{x}_\sigma) \leq n_\sigma^* Q_P^\emptyset \quad (3.20)$$

where  $n_\sigma^* \equiv \prod_{i \in \sigma} (|A_i| - 1)$  and  $n_{\min}$  is the size of the smallest outcome set.

### 3.5 Binary observables and Boolean multideviations

When the observables are all binary, i.e., when  $|A_i| = 2$  for all  $i \in B$ , the multideviations have a particularly convenient interpretation. This interpretation also applies to more general event spaces when they are viewed as binary through the use of modified multideviation functions, which will be referred to as Boolean multideviations. These turn out to be important elements for the characterization of the new Bell inequalities introduced in section 7.

#### 3.5.1 Binary observables

When the outcome sets are all binary, there is only one multideviation degree of freedom per  $\sigma$ , thanks to (3.13). All multideviations can be written in terms of an arbitrarily chosen joint outcome:

$$Q_P^{\sigma,B}(\tilde{x}_\sigma) = (-1)^{|\sigma| - |\tilde{x}_\sigma \cap \tilde{1}_\sigma|} Q_P^{\sigma,B}(\tilde{1}_\sigma) \quad (3.21)$$

We can thus simplify the notation considerably:  $Q^\sigma \equiv Q_P^{\sigma,B}(\tilde{1}_\sigma)$ .

In the binary case, we can rewrite the multideviations to give them a particularly convenient interpretation:

$$Q^\sigma = \frac{1}{2^{|B|}} (2 \Pr(\text{even } \# \text{ of } \sigma \text{ outcomes are } 2) - 1) \quad (3.22)$$

where

$$\Pr(\text{even } \# \text{ of } \sigma \text{ outcomes are } 2) = \sum_{\tilde{x}_B} P(\tilde{x}_B) \delta_{|\tilde{x}_\sigma \cap \tilde{2}_\sigma| \text{ is even}} \quad (3.23)$$

When  $\sigma = \{i, j\}$ ,

$$Pr(\text{even \# of } \{i, j\} \text{ are 2}) = P(1_i 1_j) + P(2_i 2_j) \quad (3.24)$$

which is why  $Q^{\{i,j\}}$  measures the extent to which  $x_i$  and  $x_j$  are perfectly correlated.

However, when  $\sigma = \{i, j, k\}$ ,

$$Pr(\text{even in } \tilde{2}_{\{i,j,k\}}) = P(1_i 1_j 1_k) + P(2_i 2_j 1_k) + P(2_i 1_j 2_k) + P(1_i 2_j 2_k) \quad (3.25)$$

which means  $Q^{\{i,j,k\}}$  does *not* measure the extent to which  $x_i$ ,  $x_j$ , and  $x_k$  are perfectly correlated. Rather, it measures a correlation between them that cannot be gauged by combinations of pairwise correlations among them—an irreducible fact concerning all three outcomes together. Likewise for higher orders of  $\sigma$ .

### 3.5.2 Inequality constraints

The inequality constraints deriving from  $P(\tilde{x}_B) \geq 0$  take a simple form when all observables are binary:

$$\sum_{\sigma \in \mathcal{P}(B)} (-1)^{|\sigma \cap \rho|} Q^\sigma \geq 0 \quad (3.26)$$

where  $\rho \subseteq B$ .

Of the many consequences of these inequalities, the following are the most important:

$$-Q^\emptyset \leq Q^\sigma \leq Q^\emptyset \quad (3.27)$$

and

$$\begin{aligned} Q^\sigma \text{ is maximized} &\longrightarrow \forall_{\rho \subseteq B} [Q^\rho = Q^{\sigma \ominus \rho}] \\ Q^\sigma \text{ is minimized} &\longrightarrow \forall_{\rho \subseteq B} [Q^\rho = -Q^{\sigma \ominus \rho}] \end{aligned} \quad (3.28)$$

In particular, note that maximizing  $Q^{\{i,j\}}$  causes  $Q^{\{i\}} = Q^{\{j\}}$  (perfect correlation), and minimizing  $Q^{\{i,j\}}$  causes  $Q^{\{i\}} = -Q^{\{j\}}$  (perfect anti-correlation). This is the formal expression of the geometric relation noted in the discussion of figure 2 in section 2.1.

While the behavior of the second-order multideviation is straightforward, that of the higher-orders is more subtle. One can see clearly from (3.28)

that maximizing  $Q^{\{i,j,k\}}$  will *not* cause  $Q^{\{i\}} = Q^{\{j\}} = Q^{\{k\}}$ . That would be achieved by maximizing  $Q^{\{i,j\}}$ ,  $Q^{\{i,k\}}$ , and  $Q^{\{j,k\}}$  separately. Instead,  $Q^{\{i,j,k\}}$  measures how the outcomes are correlated in a way that cannot be measured by a combination of the lower order multideviations.

The even/odd interpretation provided in the previous section elucidates this further. For example, if  $Q^\sigma$  is maximized, then the set of outcomes for observers  $\sigma$  must have an even number of 2's, and thus the extent to which observers  $\rho$  and  $\sigma \ominus \rho$  have even numbers of 2's must be correlated. If one definitely does not, then the other must not either, and vice-versa.

### 3.5.3 Boolean multideviations

Event spaces where  $|A_i| > 2$  for at least one  $i \in B$  can be viewed as binary through the use of multideviations over lattice elements, or what I will refer to as *Boolean multideviations*. The seed functions for these quantities are

$$w_B^\sigma(\tilde{x}_\sigma, \tilde{\alpha}_\sigma) \equiv \frac{1}{2^{|B|}} \prod_{i \in \sigma} (2\delta_{x_i \in \alpha_i} - 1) \quad (3.29)$$

where  $\tilde{\alpha}_\sigma \in \prod_{i \in \sigma} \mathcal{L}(A_i)$  is an intuple over the Boolean lattices of the outcome sets.

The Boolean multideviations are

$$W_f^{\sigma,B}(\tilde{\alpha}_\sigma) \equiv \sum_{\tilde{y}_B} f(\tilde{y}_B) w_B^\sigma(\tilde{y}_\sigma, \tilde{\alpha}_\sigma) \quad (3.30)$$

Although the derivation is not straightforward, the following inequalities hold:

$$\sum_{\sigma \in \mathcal{P}(B)} (-1)^{|\sigma \cap \rho|} W^\sigma \geq 0 \quad (3.31)$$

for all  $\rho \subseteq B$ . Since this is identical in form to (3.26), all of the results in sections 3.5.2 and 3.5.1 can be transferred to the general case by substituting  $Q \rightarrow W$ ,  $1_i \rightarrow \alpha_i$ , and  $2_i \rightarrow \alpha_i^c$ .

## 4 Multiple-context event spaces and distributions

### 4.1 Definition

The distributions used in Bell's theorem are not ordinary probability distributions, but rather collections of ordinary distributions, one for each possible



joint measurement context. The notational framework of multiple-context event spaces allows efficient characterization of such distributions for arbitrarily complex physical scenarios.<sup>6</sup>

A multiple-context event space is an ordered triple  $(V, M_V, N_{\cup M_V})$  representing

1. A set of observers:  $V = \{A, B, C, \dots\}$ ,
2. A set of observables for each observer  $i \in V$ :  $M_i = \{\alpha_i, \beta_i, \dots\}$ ,
3. A set of outcomes for each observable  $p_i \in M_i$ :  $N_{p_i} = \{1_{p_i}, 2_{p_i}, \dots\}$ .

Note that the indices in each set do not count the elements of the set, but rather indicate membership in it. For example, when  $p_i$  iterates over the elements of  $M_i$ ,  $i$  is held fixed and is preserved on the variable  $p_i$  primarily to emphasize that it represents an element of  $M_i$ . Likewise for  $p_i$  in the outcome  $1_{p_i}$  or outcome variable  $x_{p_i}$ .

Several spaces of physical importance are defined by the event space:

1. Joint measurement context space :  $\Pi M_V \equiv \prod_{i \in V} M_i$ ,
2. Joint outcome space:  $\Pi N_{\tilde{p}} \equiv \prod_{i \in V} N_{p_i}$ ,
3. Omni-joint outcome space:  $\Pi N_{\cup M} \equiv \prod_{i \in V} \prod_{p_i \in M_i} N_{p_i}$ .

A multiple-context probability distribution is a collection of ordinary probability distributions, one for each joint measurement context:

$$P : \Pi M_V \rightarrow \{f : \Pi N_{\tilde{q}} \rightarrow [0, 1] \mid \tilde{q} \in \Pi M_V\} \quad (4.1)$$

where

$$P_{\tilde{p}} : \Pi N_{\tilde{p}} \rightarrow [0, 1] \quad (4.2)$$

and

$$P_{\tilde{p}}(\tilde{x}_{\tilde{p}}) \geq 0 \quad (4.3)$$

$$\sum_{\tilde{x}_{\tilde{p}}} P_{\tilde{p}}(\tilde{x}_{\tilde{p}}) = 1 \quad (4.4)$$

$P_{\tilde{p}}(\tilde{x}_{\tilde{p}})$  is the probability for getting outcome  $\tilde{x}_{\tilde{p}}$  in joint measurement context  $\tilde{p}$ .

---

<sup>6</sup>See Fogel (2011) for a systematic discussion of multiple-context event spaces and distributions.

## 4.2 Multiple-context multideviations; motivating conditions

Since multideviations are defined for ordinary probability distributions, there will be a separate set of multideviations for each joint measurement context:  $Q_{P_{\tilde{p}}}^{\tilde{p}_\sigma, \tilde{p}}(\tilde{x}_{\tilde{p}_\sigma})$ . Note that  $\tilde{p}$  is being used in several different ways in this expression:  $P_{\tilde{p}}$  identifies the ordinary probability distribution to be summed over (the  $f$  in eq. 3.11),  $\tilde{p}$  on its own is the set of possible observables (the  $B$  in eq. 3.11), and  $\tilde{p}_\sigma$  is the set of observables being correlated (the  $\sigma$  in eq. 3.11).

The most common motivating conditions used in Bell's theorems have relatively simple expression in terms of multideviations.

Parameter independence (no-signalling) is a condition on distributions between different joint measurement contexts:

$$\forall \rho \ni \sigma \left[ n_{\tilde{p}_{V \setminus \sigma}} Q_{P_{\tilde{p}}}^{\tilde{p}_\sigma, \tilde{p}}(\tilde{x}_{\tilde{p}_\sigma}) = n_{\tilde{q}_{V \setminus \sigma}} Q_{P_{(\tilde{p}_\sigma \tilde{q}_{V \setminus \sigma})}}^{\tilde{p}_\sigma, (\tilde{p}_\sigma \tilde{q}_{V \setminus \sigma})}(\tilde{x}_{\tilde{p}_\sigma}) \right] \quad (4.5)$$

That is, if two joint measurement contexts share a set of observables, then the multideviations for that set will be fixed by one another.

Determinism is a condition on ordinary probability distributions. It requires the probability for one outcome to be 1 and the rest to be 0. For example, suppose  $P(\tilde{x}_B) = \delta_{\tilde{x}_B = \tilde{y}_B}$ . Then

$$Q_P^{\sigma, B}(\tilde{x}_B) = q_B^\sigma(\tilde{x}_B, \tilde{y}_B) \quad (4.6)$$

Finally, outcome independence is also a condition on ordinary distributions:

$$n_B Q_P^{\sigma, B}(\tilde{x}_\sigma) = \prod_{i \in \sigma} \left( n_B Q_P^{\{i\}, B}(x_i) \right) \quad (4.7)$$

## 4.3 The CHSH inequality

The first clue to the importance of multideviations for Bell's Theorem comes from the CHSH inequality, the simplest known tight Bell inequality. The CHSH inequality (Clauser and Horne 1974) is the strongest possible Bell-type inequality for a multiple-context distribution over a 2x2x2 event space ( $|V| = 2$ ,  $|M_i| = 2$ , and  $|N_{p_i}| = 2$ ). In the natural representation, it has the

following form:

$$P_{p_i p_j}(x_{p_i} x_{p_j}) + P_{p_i q_j}(x_{p_i} y_{q_j}) + P_{q_i p_j}(y_{q_i} x_{p_j}) - P_{q_i q_j}(y_{q_i} y_{q_j}) - P_{p_i}^{\{i\}}(x_{p_i}) - P_{p_j}^{\{j\}}(x_{p_j}) \leq 0 \quad (4.8)$$

Eight distinct inequalities of this form can be generated by choosing different values for  $\tilde{p}_V$ ,  $\tilde{q}_V$ ,  $\tilde{x}_{\tilde{p}}$ , and  $\tilde{y}_{\tilde{q}}$ . The inequality is usually expressed in the terms of arbitrarily chosen expectation values intended to simplify the appearance.

However, in the multideviation representation, the CHSH inequality has a natural simplicity:

$$-\frac{1}{2} \leq Q^{p_i p_j} + Q^{p_i q_j} + Q^{q_i p_j} - Q^{q_i q_j} \leq \frac{1}{2} \quad (4.9)$$

where  $Q^{\tilde{\mu}} = Q_{P_{\tilde{\mu}}}^{\tilde{\mu}}(\tilde{1}_{\tilde{\mu}})$ . The multideviation representation exposes an interesting fact—the constraint depends only on the second-order multideviation degrees of freedom. The appearance of the marginal degrees of freedom in the natural representation is an artifact of the representation.

The interpretation of the binary multideviations given in section 3.5.1 allows a useful understanding of the CHSH inequality. Since the binary multideviations measure how even or odd the statistics are (relative to a chosen set of outcomes), the inequality represents a limit on how incompatible the even or odd statistics for the joint measurement contexts can be with one another. For example, if we select  $\tilde{1}_{\{1,2,3,4\}}$  as a reference, then a joint outcome  $\tilde{x}_{\tilde{\mu}}$  is even or odd depending on whether an even or odd number of the outcomes are 2. Given the starting conditions, it is not possible for the joint outcomes to be odd with certainty in an odd number of the joint measurement contexts (see fig. 6); the CHSH inequalities express this fact and indicate how close the statistics generated from those conditions can get to such a state.

In section 7.2.2, the CHSH inequality is shown to be a special case of a larger classes of inequalities. The conceptual interpretation offered here is presented systematically in section 8.

## 5 A projection theorem

I will now use the multideviation framework to prove an important result: the Bell distributions—those multiple-context probability distributions satisfying all Bell inequalities—can be generated from ordinary probability distributions over the set of omni-joint outcomes. The existence of this relationship

|                   |       | Observables for B    |                      |
|-------------------|-------|----------------------|----------------------|
|                   |       | $p_B$                | $q_B$                |
| Observables for A | $p_A$ | $P(\text{Even}) = 1$ | $P(\text{Even}) = 1$ |
|                   | $q_A$ | $P(\text{Even}) = 1$ | $P(\text{Odd}) = 1$  |

Even outcomes: 11, 22  
Odd outcomes: 12, 21

Figure 6: Interpretation of the CHSH inequality. The above state is impossible for a system obeying the upper bound of equation (4.9). The state in which all of the above probabilities are zero is disallowed by the lower bound of (4.9).

is not surprising in and of itself; the equivalence of “existence of the joints” and “satisfaction of the Bell inequalities” is well-known (see fn. 7 below). The novel result is that the mapping is accomplished by ignoring specific multi-deviation degrees of freedom, namely, those involving two or more mutually exclusive observables.

## 5.1 The theorem

In the following, the event space is  $(V, M_V, N_{\cup M_V})$  (see section 4).

Deterministic distributions are those whose values are all 0 or 1. Parameter-independent (i.e., no-signalling) distributions are those whose marginals are independent of the measurement choices of other observers (see section 4.2). The deterministic, parameter-independent multiple-context distributions are thus given by

$$G_{\tilde{p}}^{\tilde{\gamma}}(\tilde{x}_{\tilde{p}}) = \delta_{\tilde{x}_{\tilde{p}} = \tilde{\gamma}_{\tilde{p}}} \quad (5.1)$$

where  $\tilde{\gamma} \in \Pi N_{\cup M}$ .

The set of distributions that satisfy the Bell inequalities, or Bell distributions, is the set of convex combinations of deterministic, parameter-independent distributions:

**Definition** (Bell distribution). All Bell distributions can be written as

$$P_{\tilde{p}}(\tilde{x}_{\tilde{p}}) = \sum_{\tilde{\gamma} \in \Pi N_{\cup M}} \mu(\tilde{\gamma}) G_{\tilde{p}}^{\tilde{\gamma}}(\tilde{x}_{\tilde{p}}) \quad (5.2)$$

where  $\mu(\tilde{\gamma}) \geq 0$  and  $\sum_{\tilde{\gamma}} \mu(\tilde{\gamma}) = 1$ .

**Theorem 1.** *Bell distributions are projections of omni-joint distributions, where the ignored degrees of freedom are multideviations involving two or more mutually-exclusive observables.*

*Proof.* First, we express an arbitrary Bell distribution in terms of omni-joint multideviations:

$$P_{\tilde{p}_V}(\tilde{x}_{\tilde{p}_V}) = \sum_{\tilde{\gamma}_{\cup M}} \mu(\tilde{\gamma}) G_{\tilde{p}}^{\tilde{\gamma}}(\tilde{x}_{\tilde{p}}) \quad (5.3)$$

$$= \sum_{\tilde{\gamma}_{\cup M}} \mu(\tilde{\gamma}) \delta_{\tilde{x}_{\tilde{p}} = \tilde{\gamma}_{\tilde{p}}} \quad (5.4)$$

$$= \sum_{\tilde{\gamma}_{\cup M}} \left( \sum_{\rho \in \mathcal{P}(\cup M)} Q_{\mu}^{\rho, \cup M}(\tilde{\gamma}_{\rho}) \right) \delta_{\tilde{x}_{\tilde{p}} = \tilde{\gamma}_{\tilde{p}}} \quad (5.5)$$

$$= \sum_{\tilde{\gamma}_{\tilde{p}}} \sum_{\tilde{\gamma}_{\cup M \setminus \tilde{p}}} \left( \sum_{\rho \in \mathcal{P}(\cup M)} Q_{\mu}^{\rho, \cup M}(\tilde{\gamma}_{\rho}) \right) \delta_{\tilde{x}_{\tilde{p}} = \tilde{\gamma}_{\tilde{p}}} \quad (5.6)$$

$$= \sum_{\tilde{\gamma}_{\tilde{p}}} \delta_{\tilde{x}_{\tilde{p}} = \tilde{\gamma}_{\tilde{p}}} \sum_{\rho \in \mathcal{P}(\cup M)} \sum_{\tilde{\gamma}_{\cup M \setminus \tilde{p}}} Q_{\mu}^{\rho, \cup M}(\tilde{\gamma}_{\rho}) \quad (5.7)$$

$$= \sum_{\tilde{\gamma}_{\tilde{p}}} \delta_{\tilde{x}_{\tilde{p}} = \tilde{\gamma}_{\tilde{p}}} \sum_{\rho \in \mathcal{P}(\cup M)} \delta_{\rho \subseteq \tilde{p}} n_{\cup M \setminus \rho} Q_{\mu}^{\rho, \cup M}(\tilde{\gamma}_{\rho}) \quad (5.8)$$

$$= \sum_{\tilde{\gamma}_{\tilde{p}}} \delta_{\tilde{x}_{\tilde{p}} = \tilde{\gamma}_{\tilde{p}}} \sum_{\rho \in \mathcal{P}(\tilde{p})} n_{\cup M \setminus \rho} Q_{\mu}^{\rho, \cup M}(\tilde{\gamma}_{\rho}) \quad (5.9)$$

$$= \sum_{\rho \in \mathcal{P}(\tilde{p})} n_{\cup M \setminus \rho} Q_{\mu}^{\rho, \cup M}(\tilde{x}_{\rho}) \quad (5.10)$$

$$= \sum_{\sigma \in \mathcal{P}(V)} n_{\cup M \setminus \tilde{p}_{\sigma}} Q_{\mu}^{\tilde{p}_{\sigma}, \cup M}(\tilde{x}_{\tilde{p}_{\sigma}}) \quad (5.11)$$

Next, we express the same distribution in terms of multideviations over the indicated joint measurement context:

$$P_{\tilde{p}_V}(\tilde{x}_{\tilde{p}_V}) = \sum_{\sigma \in \mathcal{P}(V)} Q_{P_{\tilde{p}}}^{\tilde{p}_{\sigma}, \tilde{p}}(\tilde{x}_{\tilde{p}_{\sigma}}) \quad (5.12)$$

By (3.6), for any  $\mu \subseteq V$ ,

$$\frac{n_{\tilde{p}}}{n_{\cup M}} q_{\tilde{p}}^{\tilde{p}_\rho}(\tilde{x}_{\tilde{p}_\rho}, \tilde{y}_{\tilde{p}_\rho}) = q_{\cup M}^{\tilde{p}_\rho}(\tilde{x}_{\tilde{p}_\rho}, \tilde{y}_{\tilde{p}_\rho}) \quad (5.13)$$

This means that, for any function  $f(\tilde{x}_{\tilde{p}})$ ,

$$\sum_{\tilde{x}_{\tilde{p}}} q_{\tilde{p}}^{\tilde{p}_\rho}(\tilde{x}_{\tilde{p}_\rho}, \tilde{y}_{\tilde{p}_\rho}) f(\tilde{x}_{\tilde{p}}) = \sum_{\tilde{x}_{\cup M}} q_{\cup M}^{\tilde{p}_\rho}(\tilde{x}_{\tilde{p}_\rho}, \tilde{y}_{\tilde{p}_\rho}) f(\tilde{x}_{\tilde{p}}) \quad (5.14)$$

Using (3.14) and (5.14) to manipulate (5.11) and (5.12), we derive:

$$Q_{P_{\tilde{p}}}^{\tilde{p}_\sigma, \tilde{p}}(\tilde{x}_{\tilde{p}_\sigma}) = n_{\cup M \setminus \tilde{p}_\sigma} Q_{\mu}^{\tilde{p}_\sigma, \cup M}(\tilde{x}_{\tilde{p}_\sigma}) \quad (5.15)$$

That is, given an omni-joint distribution  $\mu(\tilde{\gamma})$ , we can construct a multiple-context distribution that satisfies the Bell inequalities using only those multideviation degrees of freedom that are subsets of the joint measurement contexts,  $\tilde{p} \in \Pi M_V$ . These are just the subsets of  $\cup M$  that include no more than one element of  $M_i$  for each observer  $i \in V$ .  $\square$

## 5.2 Geometric interpretation

Theorem 1 has a straightforward geometric interpretation. The set of omni-joint distributions corresponds to a simplex in  $\mathbb{R}^{(n_{\cup M}-1)}$ . The multideviations identify orthogonal subspaces of that vector space. The theorem says that the set of Bell distributions corresponds to a polytope formed by projecting the simplex into the subspace generated by the set of multideviation vectors associated with those sets of observables that can be measured simultaneously (an affine transformation is also needed). See Appendix B for more on the geometric approach.

## 5.3 Philosophical consequences

Let

$$\Psi \equiv \{\rho \subseteq \cup M \mid \forall i \in V [ |M_i \cap \rho| \leq 1 ]\} \quad (5.16)$$

be the collection of all sets of comeasurable observables (i.e., no two are mutually exclusive).

Now imagine a world in which all observables  $\cup M$  are measured together, producing a probability distribution,  $\mu$ , over the omni-joint outcomes. Imagine, however, that the observers are not permitted to share their results with

one another; rather, some administrator takes their results, calculates the multideviations, and returns only those corresponding to elements of  $\Psi$ . The researchers will be able to recover some probability distributions corresponding to various joint measurements, but they will not be able to reconstruct  $\mu$  in total.

Theorem 1 tells us that this is effectively the situation with Bell distributions. Each Bell distribution is equivalent to at least one probability distribution over all observables taken together, but where multideviation correlations involving mutually exclusive observables are considered inaccessible.

As noted above, this result is an extension of the well-known “existence of the joints” theorem—if a multiple-context distribution satisfies the Bell inequalities, then there is a joint distribution over all observables (an “omni-joint” distribution) that reproduces the original distribution as marginals.<sup>7</sup> The novelty here is that the multideviations make clear precisely which aspects of the omni-joint distribution are hidden; or, put another way, the multideviations show us what information needs to be restored in order to reconstruct the omni-joint distribution. The inequalities (3.19) determine the ranges of allowed values for the hidden multideviations. It is when, and only when, these inequalities are inconsistent, given the observable multideviation degrees of freedom, that at least one Bell inequality is violated.

Thus, the multideviations for the elements in  $\cup M \setminus \Psi$  are, effectively, the hidden variables compatible with the Bell inequalities. A theory may have a richer set of hidden variables, but they will have to reduce to the multideviations involving mutually exclusive correlations, if the observable statistics are the Bell distributions.

## Part II

# Tight Bell inequalities

In this part, I use the projection theorem of section 5, along with matroid theory, to outline a new method for finding tight Bell inequalities. I then

---

<sup>7</sup>Fine (1982) proved this for the simplest case; Fogel (2011) proved this for the general case. The importance of this theorem has been critiqued by Svetlichny et al. (1988), Butterfield (1992), and Muller and Placek (2001).

present a new class of such inequalities, which turn out to have a straightforward interpretation, and show that they are violated by quantum mechanics.

## 6 Method for finding BIs

### 6.1 Preliminaries

A Bell inequality is an inequality satisfied by a multiple-context distribution subject to certain constraints. Some arbitrariness exists in the choice of constraints, since different choices produce the same set of distributions (and hence the same inequalities). For convenience, I have chosen to work with parameter independence (i.e., no-signalling) and determinism.<sup>8</sup>

Recall that a tight Bell inequality is an extremal, maximally restrictive Bell inequality (see fn. 2). Thus, the set of all Bell inequalities can be characterized by the complete set of tight Bell inequalities.

Geometrically, the set of Bell distributions (convex combinations of parameter-independent, deterministic distributions) corresponds to a particular polytope. The tight Bell inequalities correspond to the facets of that polytope. Thus, geometric tools can aid in the search for these inequalities.<sup>9</sup> However, much of the geometric structure involved in characterizing polytopes is irrelevant to the search for the Bell inequalities. In the following, I will use a more abstract mathematical object, the matroid, to isolate the structure important for the task at hand.

### 6.2 Matroids; duality theorem

Matroids are mathematical objects that can be used to encode the combinatoric aspects of geometric structures.<sup>10</sup> My use of matroid theory is relatively limited, so the details will be kept to a minimum here. The reader can skip to section 6.3 without significant loss of comprehension, if desired.

---

<sup>8</sup>By taking the observable distributions to be convex combinations of parameter-independent, deterministic distributions, I have also implicitly assumed that no backwards causation occurs.

<sup>9</sup>Pitowsky (1989, 1991) are the classic texts introducing geometric methods to the study of Bell inequalities.

<sup>10</sup>Matroids have a multitude of other uses, particularly in graph theory. For background on matroids, see Oxley (2011) and Björner et al. (1999).



Matroids come in two varieties, oriented and unoriented. The extra structure provided by oriented matroids is needed here, so all matroids described below should be understood as oriented.

A matroid can be characterized in a variety of ways. One way is to begin with a base set,  $E$ , and then specify a set of *bases* that satisfy a particular set of axioms. Various sets can then be defined, which carry names drawn from linear algebra and graph theory: *independent sets*, *dependent sets*, *hyperplanes*, *circuits*, *etc.*

Two types of matroids will be useful here. *Vector matroids* encode the linear independence properties of a set of vectors. *Affine matroids* encode the affine dependencies of a set of points. Any set of vectors defines a vector matroid, and any set of points, including the vertices of a polytope, defines an affine matroid.<sup>11</sup>

The task of finding the facets of a polytope is equivalent to that of finding the positive hyperplanes of the corresponding affine matroid.

Matroids have duals, which are also matroids. The bases of a dual matroid are the complements of the bases of the original.

Given a factorizable set  $\Pi A_B$ , and some subset  $\Psi \subseteq \mathcal{P}(B)$ , one can construct a polytope by projecting the simplex in  $\mathbb{R}^{|\Pi A_B|-1}$  defined by  $\Pi A_B$  into the subspace defined by the collection of multideviation vectors corresponding to the elements of  $\Psi$ . One can then use this polytope to define an affine matroid,  $M_A(\Psi)$ . After choosing an origin, one can define a vector matroid,  $M_V(\Psi)$ , using the vectors pointing from the origin to the vertices.

I have been able to prove the following result concerning these matroids:

**Theorem 2** (Duality of multideviation projections). *The affine matroid of a multideviation projection  $\Psi$  is the dual of the vector matroid of the complement  $\mathcal{P}(B) \setminus \Psi$  formed by taking the center of the polytope as the origin:*

$$M_A^*(\Psi) = M_V(\mathcal{P}(B) \setminus (\{\emptyset\} \cup \Psi)) \quad (6.1)$$

The importance of this theorem for the task at hand cannot be understated. Vector matroids are generally easier to work with than affine matroids. Furthermore, in a well-known result in matroid theory, the hyperplanes of a matroid are the complements of the circuits of the dual, and circuits are generally easier to specify than hyperplanes.

---

<sup>11</sup>The bases of vector matroid are the maximal sets of linearly independent vectors, and similarly with affine matroids.

So, the task of finding the tight Bell inequalities corresponds to that of finding the positive circuits of the vector matroid defined by the complement of  $\Psi$ , which is the set of observable multideviation degrees of freedom.

### 6.3 Necessary and sufficient conditions (TBIC)

Drawing on the duality theorem, the necessary and sufficient conditions for a set  $\Gamma \subseteq \Pi N_{\cup M}$  to define a tight Bell inequality over an event space  $(V, M, N)$  are the following:

1. There is a function  $f$  such that:

$$\begin{aligned} \text{(a)} & \forall \tilde{\gamma} \in \Pi N_{\cup M} [f(\tilde{\gamma}) = 0 \longleftrightarrow \tilde{\gamma} \notin \Gamma] \\ \text{(b)} & \forall \rho \in \mathcal{P}(\cup M) \setminus \Psi \quad \forall \tilde{\gamma} \in \Gamma [Q_f^\rho(\tilde{\gamma}_\rho) = 0] \\ \text{(c)} & \forall \tilde{\gamma} \in \Gamma [f(\tilde{\gamma}) > 0] \end{aligned} \tag{6.2}$$

where  $\Psi = \bigcup_{\tilde{p} \in \Pi M_V} \bigcup_{\sigma \in \mathcal{P}(V)} \{\tilde{p}_\sigma\}$ .

2. All functions meeting criterion 1 are scalar multiples of one another.

I will refer to these as the *tight Bell inequality conditions* (TBIC).

The connection to the above matroid result is as follows: 1a requires the function  $f$  to be a faithful representation of the set  $\Gamma$ ; 1b requires  $\Gamma$  to be a dependent set; 2 requires  $\Gamma$  to be minimally dependent (i.e., a circuit); and 1c requires  $\Gamma$  to be all positive. Thus  $\Gamma$  is a positive circuit of  $M_V(\Psi^c)$ .

Given such a set  $\Gamma$  and function  $f$ , the corresponding inequality is

$$\sum_{\tilde{\gamma}_{\cup M}} f(\tilde{\gamma}) P(\tilde{\gamma}) \geq 0 \tag{6.3}$$

which is equivalent to

$$\sum_{\sigma \in \Psi} \sum_{\tilde{\gamma}_\sigma} Q_f^\sigma(\tilde{\gamma}_\sigma) Q_P^\sigma(\tilde{\gamma}_\sigma) \geq 0 \tag{6.4}$$

Using (5.15), we can write (6.4) in terms of the multiple-context distribution (i.e., the observed statistics). What (6.4) says is that when the tight Bell inequalities are expressed in terms of multideviations of the observed statistics, the multideviations of the linear dependence function  $f$  give the coefficients.

Condition 1b of the TBIC allows a remarkable simplification, one that will prove useful below. For any  $i \in V$ ,  $p_i, q_i \in M_i$ , and  $\tilde{\gamma}_{\cup M}$ ,

$$n_{p_i} n_{q_i} f(\tilde{\gamma}) - n_{q_i} f^{\{p_i\}}(\tilde{\gamma}) - n_{p_i} f^{\{q_i\}}(\tilde{\gamma}) + f^{\{p_i, q_i\}}(\tilde{\gamma}) = 0 \quad (6.5)$$

where  $f^\sigma(\tilde{\gamma}_\sigma) \equiv \sum_{\tilde{\gamma}_{V \setminus \sigma}} f(\tilde{\gamma})$ . This form of the condition will be particularly useful in sections 6.4 and 7. (7.3) uses simplified notation to express this in even simpler form.

## 6.4 Simplification—lifting up

As Pironio (2005) showed, any tight Bell inequality specifies a similarly structured inequality for more observers, more observables, and/or different types of observables. In the next section, I will specify solutions for arbitrary numbers of observers each choosing between two binary observables. Here I will specify the “lifted up” solutions for greater numbers of observables or outcomes.

Consider an arbitrary event space,  $(V, M_V, N_{\cup M_V})$ , and a binary event space with the same number of observers,  $2_V \equiv (V, M'_V, N'_{\cup M_V})$ , where  $|M'_i| = 2$  for all  $i \in V$  and  $|N'_{p_i}| = 2$  for all  $p_i \in \cup M_V$ . Suppose  $\Gamma \subseteq \Pi N'_{\cup M'_V}$  picks out a solution of the TBIC for  $2_V$ , where  $f$  is the corresponding function. Now do the following:

1. Select some  $\tilde{p}_V, \tilde{q}_V \in \Pi M_V$  where  $p_i \neq q_i$  for all  $i \in V$ . Let  $PQ = (\tilde{p}_V \cup \tilde{q}_V)$ .
2. Relabel the elements of  $M'_V$  so that  $M'_i = \{p_i, q_i\}$  for all  $i$ .
3. Let  $L_{p_i} = \mathcal{L}(N_{p_i})$  be a boolean lattice over  $N_{p_i}$ . Select some  $\tilde{\alpha}_{PQ} \in \Pi L_{PQ}$ .<sup>12</sup>
4. Define a mapping function  $\tilde{\chi}: \Pi N_{\cup M} \rightarrow \Pi N'_{PQ}$  such that  $\forall \mu_i \in PQ$ ,

$$(\tilde{\chi}(\tilde{\gamma}))_{\mu_i} = \begin{cases} 1_{\mu_i} & \gamma_{\mu_i} \in \alpha_i \\ 2_{\mu_i} & \gamma_{\mu_i} \notin \alpha_i \end{cases} \quad (6.6)$$

5. Let  $\Gamma^* = \{\tilde{\gamma} \in \Pi N_{\cup M} \mid \tilde{\chi}(\tilde{\gamma}) \in \Gamma\}$ .

---

<sup>12</sup>The lattice intuple  $\tilde{\alpha}_{PQ}$  represents a block of intuples formed by taking the Cartesian product of the lattice elements. For example, if  $\alpha_i = \{1_i, 2_i\}$  and  $\alpha_j = \{2_j, 3_j\}$ , then  $\tilde{\alpha}_{\{i,j\}} = \{1_i 2_j, 1_i 3_j, 2_i 2_j, 2_i 3_j\}$ .

Let  $f$  be a real-valued function over  $\Pi N'_{\cup M'_V}$ . Then define  $f^*$  such that

$$f^*(\tilde{\gamma}) = f(\tilde{\chi}(\tilde{\gamma})) \quad (6.7)$$

where  $\tilde{\gamma} \in \Pi N_{\cup M}$ .

If we substitute (6.7) into (6.5) for the full event space, the resulting equations are identical to (6.5) for  $\Gamma$  in the event space  $2_V$ . Thus, if  $\Gamma$  satisfies the TBIC for  $2_V$ , then  $\Gamma^*$  must satisfy the TBIC for  $(V, M_V, N_{\cup M_V})$ .

This shows that Bell inequalities for cases where observers have many choices of arbitrarily complicated observables can be generated from Bell inequalities for cases where the same number of observers choose between two binary observables. This does *not* show that these are the only Bell inequalities for the more complicated event spaces.<sup>13</sup>

The “lifted up” solutions have essentially the same structure as the source solutions. All but a pair of observables for each observer are ignored, and the outcome space for each observable is viewed as binary (i.e., the outcome either is or is not in  $\alpha_{p_i}$ ). The resulting inequalities are thus best expressed in terms of Boolean multideviations (see section 3.5.3):

$$\sum_{\rho \in \Psi} Q_f^\rho(\tilde{\mathbf{l}}_\sigma) W_P^{\rho, \cup M}(\tilde{\alpha}_\rho) \geq 0 \quad (6.8)$$

Note that the form is invariant under changes in the number or types of observables.

## 7 Pioneer sets—new tight Bell inequalities

In this section I will present a set of solutions to the TBIC for arbitrary numbers of observers choosing between two binary observables. As shown in section 6.4, these will also generate solutions for arbitrarily complicated physical scenarios. I refer to these solutions as “pioneer sets”, for the way they way they branch out through the outcome space.

While the set of these solutions grows exponentially with the number of observers, and while solutions with genuinely new structure exist at each level, these are unfortunately but a small portion of the set of all solutions to

---

<sup>13</sup>Some inequalities for more than 2 observables per observer and more than 2 outcomes per observable that are not reducible in this way are known (see Garg and Mermin 1982; Collins and Gisin 2004).

the TBIC.<sup>14</sup> However, these solutions are relatively easily characterized and have a straightforward conceptual interpretation.

The reader uninterested in the details of these solutions can skip to section 7.2.2, where the Bell inequalities for a particularly simple subset of them are presented.

## 7.1 Definition

Given an event space  $2_V \equiv (V, M_V, N_{\cup M_V})$ , where  $M_i = \{p_i, q_i\}$  and  $N_{\mu_i} = \{1_{\mu_i}, 2_{\mu_i}\}$  for all  $i \in V$  and  $\mu_i \in M_i$ , a pioneer set is characterized by a pair  $(Z, S_Z)$ , where

1.  $Z$  is a partition of  $V$ .
2.  $S_Z$  is an indexed family of sets, where  $S_z \subseteq \mathcal{P}(z)$ , for each  $z \in Z$ .
3. For each  $z \in Z$ , and each  $i, j \in z$  where  $i \neq j$ , there is a sequence of elements of  $S_z$  such that  $i$  is in the first element,  $j$  is in the last, and every pair of consecutive elements has a non-empty intersection.<sup>15</sup>

### 7.1.1 The odd-out transformation

The odd-out transformation applies to subsets of a powerset.

Given a set  $z$  and a set  $S \subseteq \mathcal{P}(z)$ , the odd-out transformation,  $S^*$ , is

$$S^* \equiv \{\sigma \in \mathcal{P}(z) : |S \cap \mathcal{P}(\sigma)| \text{ is odd}\} \quad (7.1)$$

Note that  $S^{**} = S$ .

The odd-out transforms of the sets in the indexed family  $S_Z$  will be represented  $S_Z^*$ .

### 7.1.2 Relabeling outcomes; condition 1b of the TBIC

Each element in the omni-joint outcome space  $\Pi N_{\cup M}$  can be represented succinctly by two subsets  $\sigma, \rho \subseteq V$ :

$$(\sigma, \rho) \longleftrightarrow \tilde{1}_{\tilde{p}_{V \setminus (\sigma \ominus \rho)} \tilde{q}_{V \setminus \rho}} \tilde{2}_{\tilde{p}_{(\sigma \ominus \rho)} \tilde{q}_{\rho}} \quad (7.2)$$

---

<sup>14</sup>For example, for 3 observers, there are 352 pioneer sets. However, Pitowsky and Svozil (2001) have shown the existence of 53856 facets.

<sup>15</sup>For example, if  $z = \{A, B, C\}$ , then  $S_z = (\{A, B\}, \{B, C\})$  would satisfy this requirement, while  $S_z = (\{A, B\}, \{C\})$  would not.

With this convention, (6.5), which is equivalent to 1b of the TBIC, takes a particularly simple form:

$$f(\sigma, \rho) + f(\sigma, \rho \ominus \{i\}) = f(\sigma \ominus \{i\}, \rho) + f(\sigma \ominus \{i\}, \rho \ominus \{i\}) \quad (7.3)$$

for any  $i \in V$  and  $\sigma, \rho \subseteq V$ .

### 7.1.3 The pioneer set

Let  $\mathfrak{X}$  be the pioneer set characterized by  $(Z, S_Z)$ . Then

$$(\sigma, \rho) \in \mathfrak{X} \longleftrightarrow \forall z \in Z [(\sigma \cap z, \rho \cap z) \in \mathfrak{X}_z] \quad (7.4)$$

where

$$(\mu, \nu) \in \mathfrak{X}_z \longleftrightarrow (\mu \in S_z^* \leftrightarrow |\nu| \text{ is odd}) \quad (7.5)$$

## 7.2 The corresponding inequalities

Proof that pioneer sets define tight Bell inequalities is given in Appendix A.

### 7.2.1 General case

To specify the inequalities, which are given by (6.4) and (6.8), we need only give the multideviation of  $f$  for each element of  $\Psi$ :

$$Q_f^{\tilde{p}_{\sigma \setminus \rho} \tilde{q}_{\rho}, PQ}(\tilde{1}_{\tilde{p}_{\sigma \setminus \rho} \tilde{q}_{\rho}}) = \frac{1}{2^{|Z|}} \prod_{z \in Z} \left( \delta_{z \cap \sigma = \emptyset} + \frac{1}{2^{|z|}} \delta_{z \setminus \sigma = \emptyset} \sum_{\mu \in \mathcal{P}(z)} (-1)^{|\mu \setminus \rho|} (-1)^{\delta_{\mu \in S_z^*}} \right) \quad (7.6)$$

where  $\sigma \subseteq V$  and  $\rho \subseteq \sigma$ .

When  $|Z| > 1$ , the inequality is a straightforward composition of lower-level inequalities. In other words, inequalities for two sets of observers  $V$  and  $V'$  always define an inequality for  $V \cup V'$ .<sup>16</sup>

---

<sup>16</sup>This is likely true in general, not just for pioneer sets. Note that this is a stronger claim than that found in Pironio (2005), which only applies to cases where either  $|V| = 1$  or  $|V'| = 1$ .

### 7.2.2 Special cases; the even/odd inequalities

Genuinely new structure for each  $V$  is found when  $Z = \{V\}$ . In this case, the inequality coefficients take a simpler form:

$$Q_f^{\tilde{p}_{\sigma \setminus \rho} \tilde{q}_\rho, PQ}(\tilde{1}_{\tilde{p}_{\sigma \setminus \rho} \tilde{q}_\rho}) = \frac{1}{2} \left( \delta_{\sigma=\emptyset} + \frac{1}{2^{|V|}} \delta_{\sigma=V} \sum_{\mu \in \mathcal{P}(V)} (-1)^{|\mu \setminus \rho|} (-1)^{\delta_{\mu \in S^*}} \right) \quad (7.7)$$

Note that, aside from the constant  $Q^\emptyset$ , only the top-level multideviations ( $\sigma = V$ ) will appear in the inequality. As I will show in section 9, this has important philosophical consequences.

For reasons that will be made clear in section 8, I refer to these as *even/odd inequalities*.

From this, a series of particularly simple inequalities can be derived. For any  $\varphi \subseteq V$  and  $m \in \{0, 1\}$ ,

$$\frac{1}{2} + (-1)^m \left( 2^{|V|-1} Q_P^{\tilde{p}_{V \setminus \varphi} \tilde{q}_\varphi} - \sum_{\rho \in \mathcal{P}(V)} Q_P^{\tilde{p}_{V \setminus \rho} \tilde{q}_\rho} \right) \geq 0 \quad (7.8)$$

where  $Q_P^\mu = Q_P^\mu(\tilde{1}_\mu)$  by convention. For general event spaces, substitute  $Q_P^\mu(\tilde{1}_\mu) \rightarrow W_P^\mu(\tilde{\alpha}_\mu)$ , as in (6.8).

When  $|V| = 2$ , (7.8) reduces to the CHSH inequality (see section 4.3).

If there were no restrictions on the base states, then the  $Q_P^{\tilde{p}_{V \setminus \rho} \tilde{q}_\rho}$  could range between  $\pm \frac{1}{2^{|V|}}$  independently of one another. The left-hand side of (7.8) would then be able to reach  $-1$ , the maximal violation allowed by the probability calculus.

In section 10, I will show that quantum mechanics predicts a violation of each of these inequalities.

## 7.3 Counts

Some data on the pioneer sets is given in table 1. The total number grows roughly as  $2^{2^{|V|}}$ , and those that show genuinely new structure for a given number of observers (i.e., those where  $Z = |V|$ ) quickly come to dominate.

| $ V $ | # of pioneer sets           | # with $Z = \{V\}$          | $2^{2^{ V }}$               |
|-------|-----------------------------|-----------------------------|-----------------------------|
| 2     | 24                          | 8                           | 16                          |
| 3     | 352                         | 192                         | 256                         |
| 4     | 67,968                      | 63680                       | 65536                       |
| 5     | $\sim 4.296 \times 10^9$    | $\sim 4.294 \times 10^9$    | $\sim 4.295 \times 10^9$    |
| 6     | $\sim 1.845 \times 10^{19}$ | $\sim 1.845 \times 10^{19}$ | $\sim 1.845 \times 10^{19}$ |

Table 1: Counts of pioneer sets by number of observers.

## 8 Conceptual interpretation of the even/odd inequalities

The interpretation of the binary multideviations given in section 3.5.1 allows a straightforward interpretation of the even/odd inequalities, along the lines described in section 4.3 for the CHSH inequality. The new inequalities represent limits on how incompatible the even or odd statistics for the joint measurement contexts can be with one another.

Let a distribution be *odd-definite* if it will, with certainty, produce a joint outcome with an odd number of 2's, and *even-definite* if it will, with certainty, produce a joint outcome with an even number of 2's.<sup>17</sup> Then the even/odd inequalities express logical connections between the odd/even-definiteness of the distributions for different measurement contexts.

To see how these connections arise, consider a joint measurement of three binary observables, labeled 1, 2, and 3. If the distribution is even-definite in observables 1 and 2 and even-definite in observables 1 and 3, then it is necessarily even-definite in observables 2 and 3. For example, suppose the results of three coins being flipped are even-definite in coins 1 and 2, meaning that the joint outcome must include either  $H_1H_2$  or  $T_1T_2$ . Suppose further that the results are even-definite in 1 and 3, so that the joint outcome must include  $H_1H_3$  or  $T_1T_3$ . The total joint outcome must thus be either  $H_1H_2H_3$  or  $T_1T_2T_3$ , and the results must thus be even-definite in 2 and 3. In other words, the requirements that the results be even-definite in 12 and also in 13 imply that the results are also even-definite in 23.

We could have phrased this constraint just as easily in terms of odd-definiteness: it cannot be the case that an odd number of pairs of 1, 2, and

---

<sup>17</sup>That is, the distribution  $P_{\vec{\mu}}$  is odd- or even-definite if  $Pr_{\vec{\mu}}(\text{even \# of } V \text{ outcomes are } 2)$  is 0 or 1, respectively.





Figure 7: Even/odd compatibility for 3 observables. A vertex arrangement of  $\pm 1$  can be generated by assigning  $\pm 1$  to each side and then multiplying the adjacent values, as in the diagram on the left. Certain vertex arrangements of  $\pm 1$  cannot be constructed in this way. The diagram on the right is an example of such an arrangement.

3 are odd-definite. There will be four such constraints—one for each subset of  $\{12, 13, 23\}$  with an odd number of elements.

These limits can be represented through a simple graphical method, depicted in fig. 7. They represent possible deterministic arrangements, where the outcome of each observable is represented as  $\pm 1$ . The multideviations for pairs of observables are determined by multiplying the values for the corresponding observables. Certain arrangements among the pairwise multideviations cannot be formed in this way, namely, those where an odd number have the value  $-1$ . These represent logical restrictions on the distribution.

The even/odd inequalities are logical limitations in precisely the same way. For the 2-observer case, the corresponding graph is a square, and the impossible arrangements are also those for which an odd number of vertices have a  $-1$ . There are 8 such arrangements, corresponding to the 8 CHSH inequalities.

For more observers, the same restriction holds—the distribution cannot be odd-definite in an odd number of contexts. However, for 3 observers or more, there can also be more complicated forms of even/odd incompatibility. One such arrangement for 3 observers is depicted in fig. 8.

The graphical method generalizes straightforwardly to  $n$  observers, where the relevant graph is an  $n$ -dimensional hypercube. Each even/odd inequality corresponds to an assignment of  $\pm 1$  to the vertices that cannot be generated by assigning  $\pm 1$  to each facet and placing the product of adjacent facets at each vertex. The profile  $S$  which defines the pioneer set for the inequality is

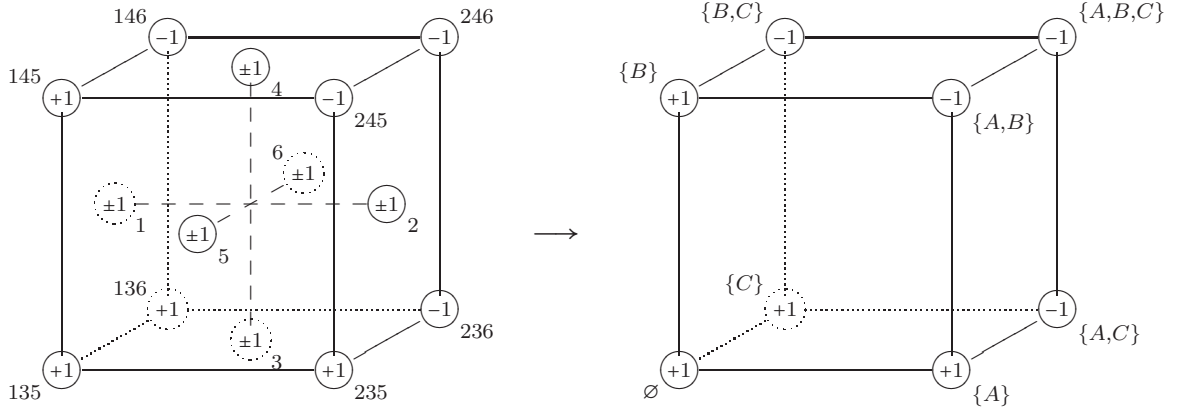


Figure 8: Example of an even-odd limitation for three observers, where  $\tilde{p} = \{1, 3, 5\}$  and  $\tilde{q} = \{2, 4, 6\}$ . The vertices represent joint measurement contexts, and the distribution is even-definite if +1 is assigned and odd-definite if -1 is assigned. There is no assignment of  $\pm 1$  to the facets of the cube such that the product of adjacent facets at each vertex produces the above arrangement. The joint measurement contexts can be labeled by subsets of  $\{A, B, C\}$ . The set of odd-definite contexts is then  $\{\{A, B\}, \{A, C\}, \{B, C\}, \{A, B, C\}\}$ , and the odd-out transform is  $\{\{A, B\}, \{A, C\}, \{B, C\}\}$ , which defines the corresponding pioneer set. The corresponding Bell inequality is given by (7.7):  $\frac{1}{4} - Q^{\{1,3,5\}} + Q^{\{1,4,6\}} + Q^{\{2,3,6\}} + Q^{\{2,4,5\}} \geq 0$ .

just the odd-out transform of the set of odd-definite contexts, when each is labeled as a subset of the set of observers (as in fig. 8).

Indeed, every inequality defined by the pioneer sets represents a kind of even/odd incompatibility, although only the even/odd inequalities can be represented so easily.

## 9 Philosophical importance of the even/odd inequalities

The even/odd inequalities are interesting not merely because they have a convenient conceptual interpretation, but also because they provide the opportunity for a stronger version of Bell's theorem. Because these inequalities concern degrees of freedom that are independent of those involved in parameter independence, they should be derivable from conditions that do not include or imply parameter independence. Violations of the even/odd inequalities could thus provide more specific demands on which classical concepts must be given up. It is not my intention here to provide such a derivation, merely to show that this is possible.

As above, I will focus on the binary case,  $2_V$ , with the proviso that all results hold for the general case under the substitution  $Q_P^\mu(\tilde{1}_\mu) \rightarrow W_P^\mu(\tilde{\alpha}_\mu)$ , as in (6.8). In the binary case, parameter independence is

$$\forall \rho \ni \sigma \left[ Q_{\tilde{p}}^\sigma = Q_{\tilde{p}\rho\tilde{q}_{V\setminus\rho}}^\sigma \right] \quad (9.1)$$

and outcome independence is

$$Q_{\tilde{p}}^\sigma = \frac{1}{2^{|V|}} \prod_{i \in \sigma} \left( 2^{|V|} Q_{\tilde{p}}^{\{i\}} \right) \quad (9.2)$$

where  $Q_{\tilde{\mu}}^\sigma \equiv Q_{P_{\tilde{\mu}}}^{\tilde{\mu}_\sigma}(\tilde{1}_{\tilde{\mu}_\sigma})$ .

The important thing to notice is that parameter independence does not affect any of the multideviations of the form  $Q_{\tilde{\mu}}^V$  (i.e., when  $\sigma = V$ ). On the other hand, outcome independence does affect these degrees of freedom. Indeed, it is through the combination of the two that the various  $Q_{\tilde{\mu}}^V$  are related to one another; outcome independence relates  $Q_{\tilde{\mu}}^V$  to the  $Q_{\tilde{\mu}}^{\{i\}}$  within a given measurement context, and parameter independence relates the  $Q_{\tilde{\mu}}^{\{i\}}$

between measurement contexts. Fig. 3 depicts these relationships for the 2 observer case.

In the notation used in this section, the even/odd inequalities are

$$1 + \sum_{\rho \in \mathcal{P}(V)} Q_{\tilde{\rho}_{V \setminus \rho} \tilde{q}_\rho}^V \left( \sum_{\mu \in \mathcal{P}(V)} (-1)^{|\mu \setminus \rho|} (-1)^{\delta_{\mu \in S^*}} \right) \geq 0 \quad (9.3)$$

where  $S^*$  is the odd-out transform of the profile set  $S \subseteq \mathcal{P}(V)$  (see 7.1). It is clear that the inequalities concern only multideviations of the form  $Q_\mu^V$ . Thus, violations of the inequalities have nothing directly to do with parameter independence. They test a type of locality (or other classical concept) that does not involve effects of the choice of measurement context.

The even/odd inequalities provide an opportunity; one could, in theory, derive them from a condition placed solely on multideviations of the form  $Q_\mu^V$ . Since these inequalities are violated by quantum mechanics (to be shown in the next section), one could then conclude that this condition cannot be satisfied by any theory aiming to reproduce quantum statistics. Since this condition would be manifestly independent of parameter independence, the result would be strictly stronger than the existing Bell's theorem beginning with parameter and outcome independence.

The trick, of course, is to find such a condition with a natural physical interpretation. Because multideviations are new quantities without well-established interpretations, there is not an obvious candidate at this time.

## 10 Quantum mechanics

I will now show that the inequalities presented in section 7.2.2 are violated by quantum mechanics. Contrary to what one might expect, the size of the violations increases with the number of observers and converges toward the theoretical maximum.

### 10.1 Experimental setup; initial state

A set of observers,  $V = \{A, B, C, \dots\}$ , each performs one of two possible spin measurements,  $M_i = \{\theta_{i,0}, \theta_{i,1}\}$ , in the  $xz$ -plane on spin- $\frac{1}{2}$  particles emitted from a central source. There are thus two possible outcomes for each measurement, and the event space has the structure of  $2_V$  (see section 7.1).

The observables  $\theta_{i,n}$  correspond to angles in the  $xz$ -plane, where 0 represents the positive direction of the  $z$ -axis and  $\frac{\pi}{2}$  represents the positive direction of the  $x$ -axis. The elements of the outcome set,  $N_{\theta_{i,n}} = \{1_{\theta_{i,n}}, 2_{\theta_{i,n}}\}$ , represent spin-up and spin-down, respectively, for the observable  $\theta_{i,n}$ .

The overall Hilbert space for the experiment is the tensor product of the 2-dimensional Hilbert spaces for each observer. States and operators will be expressed in the positive  $z$ -basis of each subspace,  $\{|1_i\rangle, |2_i\rangle\}$  for observer  $i$ .

The initial state will be prepared in an “even-correlation” state:

$$|\psi\rangle = \frac{1}{\sqrt{2^{|V|-1}}} \sum_{\substack{\sigma \in \mathcal{P}(V) \\ |\sigma| \text{ even}}} (-1)^{\frac{|\sigma|}{2}} \prod_{i \in V} (\delta_{i \notin \sigma} |1_i\rangle + \delta_{i \in \sigma} |2_i\rangle) \quad (10.1)$$

If all observers measure along the positive  $z$ -axis, then the joint outcome will always have an even number of spin-downs. As shown in the next section, this state has a remarkably simple multideviation profile, regardless of which spin-orientations are measured.

It is worth keeping in mind the difference between this state and the generalized GHZ state, which is often used to represent multi-party entanglement:

$$|\psi_{\pm}\rangle = \frac{1}{\sqrt{2}} \left( \left( \prod_{i \in V} |1_i\rangle \right) \pm \left( \prod_{i \in V} |2_i\rangle \right) \right) \quad (10.2)$$

The GHZ states are perfectly correlated in a pairwise way. If any two observers measure along the  $z$ -axis, then they will get the same result. The multideviation profile for the GHZ state is significantly more complicated than for the even-correlation state (only odd-order multideviations vanish), especially for arbitrary spin-orientations. Whether the GHZ states violate any of the top-level inequalities specified by (7.7) is unclear (for 3 or more observers).

## 10.2 Measurement results

A joint measurement context,  $\tilde{\mu} \in \Pi M_V$ , can be specified by the intuple  $\tilde{m} \in \prod_{i \in V} \{0, 1\}$ , where  $\mu_i = \theta_{i, m_i}$ . We will thus consider the measurement context to be a function of  $\tilde{m}$ :  $\tilde{\mu}(\tilde{m})$ .

The joint probabilities are

$$P_{\tilde{\mu}(\tilde{m})}(\tilde{x}) = \frac{1}{2^{|V|}} \left( 1 + (-1)^{|\tilde{x} \cap \tilde{2}|} \cos \left( \sum_{i \in V} \theta_{i, m_i} \right) \right) \quad (10.3)$$

The multideviations are

$$Q_{P_{\tilde{\mu}}}^{\emptyset, \tilde{\mu}}(\tilde{x}) = \frac{1}{2^{|V|}} \quad (10.4)$$

$$Q_{P_{\tilde{\mu}}}^{\tilde{\mu}, \tilde{\mu}}(\tilde{x}) = \frac{(-1)^{|\tilde{x} \cap \tilde{2}|}}{2^{|V|}} \cos\left(\sum_{i \in V} \theta_{i, m_i}\right) \quad (10.5)$$

where  $\tilde{\mu} = \tilde{\mu}(\tilde{m})$ . All other multideviations vanish.

Because the multideviations for different  $\tilde{x}$  differ only by (at most) a factor of -1, we will focus only on  $Q_{P_{\tilde{\mu}}}^{\sigma, \tilde{\mu}}(\tilde{1})$ .

### 10.3 Violation of the simplest Even/Odd inequalities

The simplest even/odd inequalities, given by (7.8), are indexed by  $\varphi \subseteq V$  and  $m \in \{0, 1\}$ . The state  $|\psi\rangle$  violates each one for certain choices of measurement settings.

Let

$$a_i \equiv \theta_{i, \delta_{i \in \varphi}} \quad (10.6)$$

$$d_i \equiv \frac{1}{2} (\theta_{i, \delta_{i \notin \varphi}} - a_i) \quad (10.7)$$

Then, after some algebraic manipulation, (7.8) becomes

$$\frac{1}{2} + (-1)^{|m|} \frac{1}{2} \left( \cos\left(\sum_{i \in V} a_i\right) - 2 \cos\left(\sum_{i \in V} (a_i + d_i)\right) \left(\prod_i^V \cos(d_i)\right) \right) \geq 0 \quad (10.8)$$

Now, let

$$a \equiv \left(\sum_{i \in V} a_i\right) + \pi \delta_{m=0} \quad (10.9)$$

$$d_i = \frac{\pi}{2^{|V|}} \quad (10.10)$$

and the inequality becomes

$$\frac{1}{2} - \left(\frac{1}{2} \cos a + \sin a \cos^{|V|}\left(\frac{\pi}{2^{|V|}}\right)\right) \geq 0 \quad (10.11)$$

When  $a = 0$ , the left-hand side is 0, and the inequality is thus satisfied through equality. However, the derivative with respect to  $a$  is

$$-\left(-\frac{1}{2} \sin a + \cos a \cos^{|V|}\left(\frac{\pi}{2|V|}\right)\right) \quad (10.12)$$

At  $a = 0$ , this is  $-\cos^{|V|}\left(\frac{\pi}{2|V|}\right)$ , which is manifestly negative. Thus, for values of  $a$  slightly larger than 0, the inequality is violated.

## 10.4 Maximal violations

If, instead of (10.10), we assume  $d_i = \frac{d}{|V|}$ , then we get the inequality

$$\frac{1}{2} - \left(\frac{1}{2} \cos a - \cos(a + d) \cos^{|V|}\left(\frac{d}{|V|}\right)\right) \geq 0 \quad (10.13)$$

This expression is minimized over variations in  $a$  when  $a = \pi - d\left(\frac{|V|+1}{|V|}\right)$ . Then it becomes

$$\frac{1}{2} - \left(-\frac{1}{2} \cos\left(d + \frac{d}{|V|}\right) + \cos^{|V|+1}\left(\frac{d}{|V|}\right)\right) \quad (10.14)$$

Finding the minimum is difficult analytically, but a numerical search is straightforward (see table 2).

| $ V $ | $\frac{d}{\pi}$ | Value of (10.14) |
|-------|-----------------|------------------|
| 2     | 0.5             | -0.207           |
| 3     | 0.588           | -0.333           |
| 4     | 0.689           | -0.421           |
| 5     | 0.802           | -0.487           |
| 10    | 0.972           | -0.669           |
| 100   | 0.997           | -0.953           |
| 1000  | 0.999           | -0.999           |

Table 2: Maximal violations of (10.11).

Most important, the maximal violation of the inequality increases with  $|V|$ . As  $|V| \rightarrow \infty$ , the expression converges to

$$\frac{1}{2} (\cos d - 1) \quad (10.15)$$

This is obviously minimized when  $d = \pi$ , where it is equal to  $-1$ . As noted in section 7.2.2, that is the maximal violation allowed by the probability calculus (i.e., where there are no restrictions on the underlying distributions).

## 11 Conclusion

The introduction of multideviations exposes some of the underlying structure of the distributions described by Bell’s theorem. In particular, those distributions can be generated from joint distributions over all observables by ignoring specific multideviation degrees of freedom, namely, those involving pairs of mutually exclusive observables. Thus, further study of multideviations should help illuminate the philosophical importance of Bell’s theorem.

The new method for finding tight Bell inequalities presented above does reduce the computational complexity of the problem somewhat, but not enough to keep brute force calculations from being intractable for relatively small numbers of observers. Still, the new organization of the problem may prompt further improvements.

The presentation of new tight Bell inequalities for arbitrary numbers of observers, particularly the even/odd inequalities, which have relatively simple form and admit convenient conceptual interpretation, allows for the confirmation that quantum mechanics violates the assumptions of Bell’s theorem (however they are formulated) for any number of systems. Furthermore, the size of this violation increases with the number of systems and converges to the theoretical maximum.

This last fact is somewhat surprising, for two different reasons. First, quantum effects tend to be dampened in general as the number of systems is increased. Yet, if we take the violation of a Bell inequality to indicate something peculiarly non-classical about an experiment, then the effect is *more* pronounced as the number of systems increases.

Second, the fact that quantum mechanics does not permit a maximal violation of the CHSH inequality has sparked a significant amount of interest.<sup>18</sup> The hope has been that some physical principle will explain the limitation and perhaps provide some non-empirical justification for the Schrödinger equation. The above result, which shows that violation of Bell inequalities converges toward the theoretical maximum as the number of systems is in-

---

<sup>18</sup>For some recent work, see Filipp and Svozil (2005); Janotta et al. (2011).



creased, suggests that this limitation is a peculiarity of lower-dimensional systems and perhaps of less fundamental importance than it may seem.

Finally, the even/odd inequalities concern degrees of freedom that are unaffected by parameter independence, raising the possibility of a new Bell's theorem that omits this condition altogether. Such a theorem would allow a stronger philosophical result, namely, a more precise articulation of the classical concepts that cannot be part of any empirically adequate future physics.

## References

- Bancal, J.-D., Brunner, N., Gisin, N., and Liang, Y.-C. (2011). Detecting Genuine Multipartite Quantum Nonlocality: A Simple Approach and Generalization to Arbitrary Dimensions. *Physical Review Letters*, 106(2):4.
- Bell, J. S. (1964). On the Einstein-Podolsky-Rosen paradox. *Physics*, I:195–200.
- Bell, J. S. (1966). On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.*, 38(3):447–452.
- Björner, A., Las Vergnas, M., Sturmfels, B., White, N., and Ziegler, G. (1999). *Oriented matroids*, volume 46 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, 2nd edition.
- Butterfield, J. (1992). Bell's theorem: What it takes. *British Journal for the Philosophy of Science*, 43(1):41–83.
- Clauser, J. F. and Horne, M. A. (1974). Experimental consequences of objective local theories. *Phys. Rev. D*, 10(2):526–535.
- Collins, D. and Gisin, N. (2004). A relevant two qubit Bell inequality inequivalent to the CHSH inequality. *Journal of Physics A: Mathematical and General*, 37(5):1775–1787.
- Collins, D., Gisin, N., Popescu, S., Roberts, D., and Scarani, V. (2002). Bell-type inequalities to detect true n-body nonseparability. *Physical Review Letters*, 88(17).

- Filipp, S. and Svozil, K. (2005). Tracing the bounds on Bell-type inequalities. In *AIP Conference Proceedings: Foundations of Probability and Physics - 3*, pages 87–94. AIP.
- Fine, A. (1982). Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letters*, 48(5):291–295.
- Fogel, B. (2011). Multiple-context event spaces and distributions: A new framework for Bell’s theorems. In progress.
- Garg, A. and Mermin, N. (1982). Correlation Inequalities and Hidden Variables. *Physical Review Letters*, 49:1220–1223.
- Greenberger, D., Horne, M., Shimony, A., and Zeilinger, A. (1990). Bell’s theorem without inequalities. *American Journal of Physics*, 58(12):1131–1143.
- Janotta, P., Gogolin, C., Barrett, J., and Brunner, N. (2011). Limits on nonlocal correlations from the structure of the local state space. *New Journal of Physics*, 13(6):063024.
- Mermin, N. (1990). Simple unified form for the major no-hidden-variables theorems. *Physical Review Letters*, 65(27):3373–3376.
- Muller, T. and Placek, T. (2001). Against a minimalist reading of Bell’s theorem: Lessons from Fine. *Synthese*, 128(3):343–379.
- Oxley, J. G. (2011). *Matroid theory*, volume 21 of *Oxford graduate texts in mathematics*. Oxford University Press, Oxford, 2nd ed edition.
- Peres, A. (1999). All the Bell inequalities. *Foundations of Physics*, 29(4):589–614.
- Pironio, S. (2005). Lifting Bell inequalities. *Journal of Mathematical Physics*, 46(6):062112.
- Pitowsky (1989). Quantum probability – quantum logic. Springer-Verlag.
- Pitowsky, I. (1991). Correlation polytopes: Their geometry and complexity. *Mathematical Programming*, 50:395–414.

- Pitowsky, I. and Svozil, K. (2001). Optimal tests of quantum nonlocality. *Phys. Rev. A*, (014102):4.
- Svetlichny, G. (1987). Distinguishing three-body from two-body nonseparability by a Bell-type inequality. *Physical Review D*, 35(10):3066–3069.
- Svetlichny, G., Redhead, M., Brown, H., and Butterfield, J. (1988). Do the Bell inequalities require the existence of joint probability distributions? *Philosophy of Science*, 55(3):387.
- Uffink, J. (2002). Quadratic Bell inequalities as tests for multipartite entanglement. *Physical Review Letters*, 88(230406):4.
- Werner, R. F. and Wolf, M. M. (2001). All-multipartite Bell-correlation inequalities for two dichotomic observables per site. *Phys. Rev. A*, 64(3):032112.
- Żukowski, M. and Brukner, Č. (2002). Bell’s theorem for general n-qubit states. *Physical Review Letters*, 88(210401):4.

## A Proof that pioneer sets define tight Bell inequalities

### A.1 Difference function

The following is a straightforward consequence of (7.5):

$$(\mu, \nu) \in \mathfrak{X}_z \underline{\vee} (\mu, \nu \ominus \{i\}) \in \mathfrak{X}_z \quad (\text{A.1})$$

for any  $i \in z$  and  $\mu, \nu \subseteq z$ .

This allows the definition of a kind of discrete differential:

$$\mathfrak{K}(\mu, \nu, i) \equiv \begin{cases} \emptyset & (\mu, \nu) \in \mathfrak{X}_z \leftrightarrow (\mu \ominus \{i\}, \nu) \in \mathfrak{X}_z \\ \{i\} & (\mu, \nu) \in \mathfrak{X}_z \leftrightarrow (\mu \ominus \{i\}, \nu \ominus \{i\}) \in \mathfrak{X}_z \end{cases} \quad (\text{A.2})$$

Thus,

$$(\mu, \nu) \in \mathfrak{X}_z \leftrightarrow (\mu \ominus \{i\}, \nu \ominus \mathfrak{K}(\mu, \nu, i)) \in \mathfrak{X}_z \quad (\text{A.3})$$

By (7.5), it can be shown that

$$\mathfrak{K}(\mu, \nu, i) \equiv \begin{cases} \emptyset & \mu \in S_z^* \leftrightarrow \mu \ominus \{i\} \in S_z^* \\ \{i\} & \mu \in S_z^* \not\subseteq \mu \ominus \{i\} \in S_z^* \end{cases} \quad (\text{A.4})$$

Thus,  $\mathfrak{K}(\mu, \nu, i) = \mathfrak{K}(\mu, i)$ . Note, also, that  $\mathfrak{K}(\mu, i) = \mathfrak{K}(\mu \ominus \{i\}, i)$ .

Finally, (7.3) and (A.3) combined with 1a of the TBIC mean that

$$f(\mu, \nu) = f(\mu \ominus \{i\}, \nu \ominus \mathfrak{K}(\mu, i)) \quad (\text{A.5})$$

Thus, what we need to do is show that the difference function  $\mathfrak{K}$  can connect every two elements of each  $S_z$ .

## A.2 Vertical slices

We will now construct a series of sequences where consecutive elements are related by (A.5).

Select some  $\nu$  such that  $(\emptyset, \nu) \in S_z$  and some  $\mu \subseteq V$ . Let  $\mu_n$  be the  $n$ th element of  $\mu$ . Construct a sequence of  $|\mu|$  elements according to the following:

$$(a_0, b_0) = (\emptyset, \nu) \quad (\text{A.6})$$

$$(a_n, b_n) = (a_{n-1} \ominus \{\mu_n\}, b_{n-1} \ominus \mathfrak{K}(a_{n-1}, \mu_n)) \quad (\text{A.7})$$

Then  $f(a_n, b_n) = f(a_{n-1}, b_{n-1}) = f(\emptyset, \nu)$  for all  $n$ .

By maintaining a constant ordering of the elements of  $z$ , it is simple to show that all elements of  $\mathfrak{X}_z$  are partitioned according to which element  $(\emptyset, \nu)$  they can be connected to in the above manner. Thus, we now need only show that the elements  $(\emptyset, \nu)$  can be connected to one another.

## A.3 Horizontal slices

We now want to show that, given any  $i, j \in z$ ,

$$f(\emptyset, \nu) = f(\emptyset, \nu \ominus \{i, j\}) \quad (\text{A.8})$$

We will do so by constructing a sequence connecting the two elements.

Select some  $Y \in S_z$  such that  $\{i, j\} \subseteq Y$ . By #3 in the definition of the pioneer set (see section 7.1), such a  $Y$  must exist. Now take some  $\sigma \subseteq Y \setminus \{i, j\}$  and choose an arbitrary order. Let  $\sigma_n$  be the  $n$ th element of  $\sigma$ , and let  $\sigma_{\leq n}$  be the first  $n$  elements of  $\sigma$ .

We now construct a sequence in three parts. Let  $(a_0, b_0) = (\emptyset, \nu)$ . For  $1 \leq n \leq |\sigma|$ , the sequence is similar to that used above:

$$(a_n, b_n) = (a_{n-1} \ominus \{\sigma_n\}, b_{n-1} \ominus \mathfrak{K}(a_{n-1}, \sigma_n)) \quad (\text{A.9})$$

The next 4 elements are:

$$(a_{|\sigma|+1}, b_{|\sigma|+1}) = (\sigma \ominus \{i\}, b_{|\sigma|} \ominus \mathfrak{K}(\sigma, i)) \quad (\text{A.10})$$

$$(a_{|\sigma|+2}, b_{|\sigma|+2}) = (\sigma \ominus \{i, j\}, b_{|\sigma|+1} \ominus \mathfrak{K}(\sigma \ominus \{i\}, j)) \quad (\text{A.11})$$

$$(a_{|\sigma|+3}, b_{|\sigma|+3}) = (\sigma \ominus \{j\}, b_{|\sigma|+2} \ominus \mathfrak{K}(\sigma \ominus \{i, j\}, i)) \quad (\text{A.12})$$

$$(a_{|\sigma|+4}, b_{|\sigma|+4}) = (\sigma, b_{|\sigma|+3} \ominus \mathfrak{K}(\sigma \ominus \{j\}, j)) \quad (\text{A.13})$$

The final part of the sequence, another  $|\sigma|$  elements ( $|\sigma|+5 \leq n \leq 2|\sigma|+4$ ), is the reverse of the first part:

$$(a_n, b_n) = (a_{n-1} \ominus \{\sigma_{|\sigma|-(n-(|\sigma|+5))}\}, b_{n-1} \ominus \mathfrak{K}(a_{n-1}, \sigma_{|\sigma|-(n-(|\sigma|+5))})) \quad (\text{A.14})$$

So,  $a_{2|\sigma|+4} = \emptyset$ . As before,  $f(a_n, b_n) = f(a_{n-1}, b_{n-1}) = f(\emptyset, \nu)$  for all  $n$ .

Recall that  $\mathfrak{K}(\mu, i) = \mathfrak{K}(\mu \ominus \{i\}, i)$ . This means that

$$b_{n-1} \ominus b_n = b_{(2|\sigma|+4)-n} \ominus b_{(2|\sigma|+5)-n} \quad (\text{A.15})$$

In other words, the first and last parts of the sequence cancel each other out, so that

$$b_{(2|\sigma|+4)-n} = \nu \ominus (b_{|\sigma|} \ominus b_{|\sigma|+4}) \quad (\text{A.16})$$

and

$$(b_{|\sigma|} \ominus b_{|\sigma|+4}) = \mathfrak{K}(\sigma, i) \ominus \mathfrak{K}(\sigma \ominus \{i\}, j) \ominus \mathfrak{K}(\sigma \ominus \{i, j\}, i) \ominus \mathfrak{K}(\sigma \ominus \{j\}, j) \quad (\text{A.17})$$

Let  $\tau_\sigma \equiv (b_{|\sigma|} \ominus b_{|\sigma|+4})$ . Then,

$$i \in \tau_\sigma \longleftrightarrow (i \in \mathfrak{K}(\sigma, i) \underline{\vee} i \in \mathfrak{K}(\sigma \ominus \{i, j\}, i)) \quad (\text{A.18})$$

$$\longleftrightarrow (\sigma \in S_z^* \underline{\vee} \sigma \ominus \{i\} \in S_z^* \underline{\vee} \sigma \ominus \{j\} \in S_z^* \underline{\vee} \sigma \ominus \{i, j\} \in S_z^*) \quad (\text{A.19})$$

Finally, we create a much longer sequence by joining together such sequences for all  $\sigma \subseteq Y \setminus \{i, j\}$ . The final element in this large sequence will be  $(\emptyset, \nu')$ , where

$$\nu' = \bigoplus_{\sigma \subseteq Y \setminus \{i, j\}} \tau_\sigma \quad (\text{A.20})$$

and

$$i \in \nu' \longleftrightarrow \bigvee_{\rho \in Y} \rho \in S_z^* \quad (\text{A.21})$$

$$\longleftrightarrow Y \in S_z \quad (\text{A.22})$$

Since  $Y \in S_z$  by assumption,  $i \in \nu'$ . By symmetry,  $j \in \nu'$ . Thus,

$$f(\emptyset, \nu) = f(\emptyset, \nu \ominus \{i, j\}) \quad (\text{A.23})$$

which is what we set out to show.

Since  $i, j$  were chosen arbitrarily, then for any  $\mu, \nu, \rho \subseteq z$ , where  $|\rho|$  is even,

$$f(\emptyset, \nu) = f(\mu, \nu \ominus \rho) \quad (\text{A.24})$$

## A.4 Final steps

The reasoning in the previous subsections can be repeated to show that, for any  $\sigma, \rho, \tau \subseteq V$ , where  $|\tau|$  is even,

$$f(\emptyset, \rho) = f(\sigma, \rho \ominus \tau) \quad (\text{A.25})$$

Thus, there is some scalar  $c$ , such that for all  $(\sigma, \rho) \in \mathfrak{X}$ ,

$$f(\sigma, \rho) = c \quad (\text{A.26})$$

The TBIC are thus satisfied. A nonzero solution exists in which all values are positive ( $c > 0$ ), and if any elements are removed from  $\Gamma$ , then no non-zero solution exists (because  $c$  would have to equal 0). QED.

## B Geometric epilogue

### B.1 Ordinary probability distributions

Any probability distribution can be represented as a vector in a vector space; this allows the set of distributions to be evaluated geometrically. The unconstrained set of ordinary probability distributions over a finite, discrete set of cardinality  $n$  forms an especially simple shape, known as a simplex, in  $\mathbb{R}^n$ . The multiple-context distributions of interest in Bell's Theorem form a more complicated polytope, one whose shape encodes the information in

the Bell-type inequalities. In this section, I will describe how to employ multideviations in the geometric description of probability distributions.

Given a product set  $\Pi A_B$ , there is a one-to-one correspondence between real-valued functions over that set and vectors in  $\mathbb{R}^{n_B}$ , where  $n_B$  is the cardinality function:

$$\vec{f} = \sum_{\tilde{x}_B} f(x) \hat{e}_{\tilde{x}} \quad (\text{B.1})$$

$$f(x) = \vec{f} \cdot \hat{e}_x \quad (\text{B.2})$$

where  $\{\hat{e}_{\tilde{x}_B}\}_{\tilde{x}_B}$  is an orthonormal basis.

The vectors corresponding to the set of probability distributions over  $\Pi A_B$  form a subset of  $\mathbb{R}^{n_B}$ , an  $(n_B - 1)$ -dimensional simplex, the simplest kind of convex polytope (the regular polygons and regular solids are low-dimensional convex polytopes; triangles and tetrahedra are examples of simplexes). Each vertex of the simplex corresponds to a distribution in which one outcome has a probability of 1, and each facet corresponds to an inequality requiring a particular probability to be greater than zero. Restrictions on the probability distribution will produce polytopes with more complex shapes.

## B.2 Multideviation vectors

The MSFs can be used to define multideviation vectors (MD-vectors), which carve up the vector space into orthogonal subspaces:

$$\vec{q}^\sigma(\tilde{x}_\sigma) \equiv \sum_{\tilde{y}_B} q^\sigma(\tilde{x}_\sigma, \tilde{y}_\sigma) \hat{e}_{\tilde{y}} \quad (\text{B.3})$$

The MD-vectors decompose the basis vectors:

$$\hat{e}_{\tilde{x}} = \sum_{\sigma \in \mathcal{P}(B)} \vec{q}^\sigma(\tilde{x}_\sigma) \quad (\text{B.4})$$

MD-vectors of different order are orthogonal:

$$\vec{q}^\sigma(\tilde{x}_\sigma) \cdot \vec{q}^\mu(\tilde{y}_\mu) = \delta_{\sigma=\mu} q^\sigma(\tilde{x}_\sigma, \tilde{y}_\sigma) \quad (\text{B.5})$$

The MD-vectors thus carve up the vector space into orthogonal subspaces. Within each  $\sigma$ -subspace, the MD-vectors are *not* orthogonal; indeed, they are

not even linearly independent. This is apparent in the summation property, inherited from the MSFs:

$$\forall_{i \in \sigma} \left[ \sum_{x_i} \tilde{q}^\sigma(\tilde{x}_\sigma) = 0 \right] \quad (\text{B.6})$$

On the other hand, the set of MD-vectors for a given  $\sigma$  does span the  $\sigma$ -subspace, and the set of  $\sigma$ -subspaces spans all of  $\mathbb{R}^{n_B}$ . The MD-vectors thus operate as a pseudo-basis; they can be used to decompose an arbitrary vector, and their inner product with that vector gives the size of the component:

$$\vec{v} = \sum_{\sigma \in \mathcal{P}(B)} \sum_{\tilde{x}_B} (\vec{v} \cdot \tilde{q}^\sigma(\tilde{x}_\sigma)) \tilde{q}^\sigma(\tilde{x}_\sigma) \quad (\text{B.7})$$

MD-vectors map functions to vectors:

$$\vec{f} \cdot \tilde{q}^\sigma(\tilde{x}_\sigma) = Q_f^\sigma(\tilde{x}_\sigma) \quad (\text{B.8})$$

For an ordinary probability distribution, then,

$$\vec{P} = \sum_{\sigma \in \mathcal{P}(B)} \sum_{\tilde{x}_B} \tilde{q}^\sigma(\tilde{x}_\sigma) Q_P^\sigma(\tilde{x}_\sigma) \quad (\text{B.9})$$

As noted above, the set of all such vectors is an  $(n_B - 1)$ -dimensional simplex.

### B.3 Multideviation polytopes

Because the multideviations are segregated into orthogonal subspaces by their order, we can use them to specify projections into a large group of subspaces. Such projections will generate new polytopes from the fundamental simplex. For every  $\Psi \subseteq \mathcal{P}(B)$ , the *multideviation polytope* given by

$$\vec{P}_\Psi = \sum_{\sigma \in \Psi} \sum_{\tilde{x}_B} \tilde{q}^\sigma(\tilde{x}_\sigma) Q_P^\sigma(\tilde{x}_\sigma) \quad (\text{B.10})$$

Each polytope corresponds to the set of distributions that are possible when certain correlation degrees of freedom are considered accessible. Finding the facet structure of an arbitrary multideviation polytope is likely to be an NP-hard problem.

As the projection theorem of section 5 shows, the Bell polytopes are a subclass of multideviation polytopes. In particular, the Bell polytope is a multideviation polytope where only correlations corresponding to simultaneously realizable joint measurements are considered accessible.



## B.4 Multiple-context distributions

For a geometric characterization of multiple-context distributions, a richer vector space structure is required. There needs to be a separate vector space for each joint measurement context  $\tilde{p} \in \Pi M_V$ :

$$\hat{e}_{\tilde{x}}^{\tilde{p}} \cdot \hat{e}_{\tilde{y}}^{\tilde{q}} = \delta_{\tilde{p}=\tilde{q}} \delta_{\tilde{x}=\tilde{y}} \quad (\text{B.11})$$

The vectors and distributions are related in a straightforward fashion:

$$\vec{P} = \sum_{\tilde{p}} \sum_{\tilde{x}} P_{\tilde{p}}(\tilde{x}) \hat{e}_{\tilde{x}}^{\tilde{p}} \quad (\text{B.12})$$

The distribution is recovered from the vector in an equally straightforward way:

$$P_{\tilde{p}}(\tilde{x}) = \vec{P} \cdot \hat{e}_{\tilde{x}}^{\tilde{p}} \quad (\text{B.13})$$

The set of all such vectors describes a convex polytope, although not a simplex.<sup>19</sup> Once constraints are added, a polytope with a more complicated shape arises, and the task of finding the tight Bell inequalities is equivalent to that of finding the facets of the polytope. Generating a description of a polytope in terms of its facets (known as the  $\mathcal{H}$ -representation) given a description in terms of its vertices (the  $\mathcal{V}$ -representation) is known as the convex hull problem. For arbitrary polytopes, the hull problem is known to be NP-complete (see Pitowsky 1991).

---

<sup>19</sup>It is, rather, the geometric product of a simplex for each joint measurement context.